

Análise de falhas em elementos mecânicos: aplicação de métodos fiabilísticos

José Luís Henriques da Silva

Dissertação realizada no âmbito do
Programa Doutoral em Engenharia Mecânica

Orientador: Professor Doutor Luís António de Andrade Ferreira

2016

2016

Resumo

O principal objetivo deste trabalho consiste na estimação dos parâmetros da distribuição fiabilística que melhor se ajusta aos dados provenientes de um registo histórico, de modo a conhecer o comportamento do equipamento ou do componente.

O método de máxima verosimilhança permite estimar os parâmetros desconhecidos do modelo estatístico. Estes parâmetros são obtidos através da maximização da função de verosimilhança do modelo em análise.

Em muitas situações práticas a função de verosimilhança está associada a modelos complexos e a equação de verosimilhança não apresenta solução analítica explícita, só sendo possível a sua resolução através de métodos numéricos. A estimação dos parâmetros da distribuição de Weibull pelo método de máxima verosimilhança a partir da informação proveniente de um registo histórico com dados censurados à direita apresenta essa dificuldade.

A solução apresentada neste trabalho passa pela utilização do algoritmo Expectation-Maximization (EM).

Dados reais provenientes do registo histórico de falhas de 5 bombas centrífugas de uma empresa petroquímica foram analisados para aplicação da metodologia.

Dado que o registo histórico de falhas apresentava um reduzido número de dados, a determinação do intervalo de confiança dos parâmetros estimados foi obtido pelo método bootstrap.

Palavras-chave: Algoritmo EM, estimação paramétrica, método de máxima verosimilhança, fiabilidade.

Abstract

The main objective of this study is the estimation of distribution parameters that best fits the data from a historical record in order to meet the equipment or component behavior.

The maximum-likelihood estimation (MLE) is a method of estimating the parameters of a statistical model given data. This method allows us to estimate the unknown parameters of a statistical model. These parameters are obtained by maximizing the likelihood function of the model in question.

In many practical situations the likelihood function is associated with complex models and the likelihood equation has no explicit analytical solution, it is only possible to have its resolution through numerical methods. The estimation of the parameters of the Weibull distribution by maximum-likelihood method based on information from a historical record with data censored on the right shows this difficulty.

The solution presented in this work entails using the Expectation-Maximization (EM) algorithm.

Actual data from the historical record of 5 centrifugal pumps failures of a petrochemical company were analyzed for application of the methodology.

Keywords: *EM algorithm, parameter estimation, maximum likelihood estimates, reliability*

Agradecimentos

Em primeiro lugar quero expressar os mais profundos agradecimentos ao meu orientador, o Professor Doutor Luís António de Andrade Ferreira, pela orientação, ajuda e amizade prestada durante a elaboração desta dissertação.

Agradeço ao Professor Daniel Gaspar pela sua ajuda e motivação para o desenvolvimento deste trabalho, a quem ficarei eternamente agradecido.

Agradeço igualmente à Professora Gabriela Direito pela sua disponibilidade e ajuda demonstrada.

Aos meus colegas do Departamento de Engenharia Mecânica e Gestão Industrial da Escola Superior de Tecnologia do Politécnico de Viseu, em especial, ao Doutor Alexandre Aibeo e ao Doutor Paulo Vaz, pelo apoio sempre demonstrado.

O meu agradecimento à Galp Energia, nomeadamente ao Engenheiro Carlos Fagundes e ao Engenheiro António Freitas por me facultarem os elementos referentes aos equipamentos estudados.

À minha esposa, ao meu filho, aos meus pais e à minha irmã que estiveram sempre presentes e me apoiaram nos momentos mais difíceis, pela paciência que sempre tiveram, pelo incentivo que sempre manifestaram.

A todos os amigos mencionados e a todos que não o foram, mas que de algum modo contribuíram para que eu pudesse realizar este trabalho, os meus sinceros e profundos agradecimentos.

Índice

| | |
|--|-------|
| Resumo | iii |
| Abstract | v |
| Agradecimentos | vii |
| Índice | ix |
| Lista de figuras | xiii |
| Lista de tabelas | xvi |
| Abreviaturas e símbolos | xviii |
| Capítulo 1 - Introdução | 1 |
| 1.1 - Estrutura da tese | 2 |
| Capítulo 2 - Metodologia de estimação paramétrica | 5 |
| 2.1 - Introdução | 5 |
| 2.2 - Enquadramento | 5 |
| 2.3 - Metodologia de estimação paramétrica | 11 |
| 2.3.1 - Passo 1: Recolha de dados | 11 |
| 2.3.2 - Passo 2: Análise preliminar dos dados | 12 |
| 2.3.2.1 - Dados censurados à direita | 13 |
| 2.3.2.1.1 - Dados censurados à direita Tipo 1 | 14 |
| 2.3.2.1.2 - Dados censurados à direita Tipo 2 | 14 |
| 2.3.2.2 - Dados censurados à esquerda | 14 |
| 2.3.2.3 - Dados censurados por intervalo | 14 |
| 2.3.2.4 - Dados truncados | 16 |
| 2.3.3 - Passo 3: Seleção da distribuição | 18 |

| | |
|---|-----------|
| 2.3.4 - Passo 4: Estimação dos parâmetros da distribuição | 19 |
| 2.3.5 - Passo 5: Validação dos resultados | 21 |
| 2.4 - Conclusões | 22 |
| Capítulo 3 - Seleção da distribuição | 23 |
| 3.1 - Introdução | 23 |
| 3.2 - Fiabilidade | 23 |
| 3.3 - Sistemas não reparáveis | 26 |
| 3.4 - Sistemas reparáveis | 28 |
| 3.4.1 - Processo de renovação | 32 |
| 3.4.2 - Processo de Poisson homogéneo | 33 |
| 3.4.3 - Processo de Poisson não homogéneo | 34 |
| 3.5 - Análise de tendência | 36 |
| 3.5.1 - Métodos gráficos | 36 |
| 3.5.1.1 - Gráfico do número acumulado de falhas | 36 |
| 3.5.1.2 - Gráfico de Nelson-Aalen | 37 |
| 3.5.1.3 - Gráfico TTT (tempo total em teste) | 38 |
| 3.5.2 - Métodos analíticos | 39 |
| 3.5.2.1 - Teste de Mann | 39 |
| 3.5.2.2 - Teste de Laplace | 39 |
| 3.5.2.3 - Teste de Lewis-Robinson | 40 |
| 3.5.2.4 - Teste MIL-HDBK | 41 |
| 3.6 - Modelo de processo de seleção | 41 |
| 3.7 - Caracterização da distribuição selecionada | 43 |
| 3.7.1 - Distribuição de Weibull | 43 |
| 3.8 - Conclusões | 48 |
| Capítulo 4 - Estimação dos parâmetros da distribuição de Weibull | 49 |
| 4.1 - Introdução | 49 |
| 4.2 - Métodos de estimação dos parâmetros da distribuição de Weibull | 49 |
| 4.2.1 - Método gráfico de estimação | 50 |
| 4.2.1.1 - Estimação de $F(t)$ | 51 |

| | |
|---|------------|
| 4.2.2 - Métodos analíticos de estimação | 55 |
| 4.2.2.1 - Método de estimação de mínimos quadrados | 55 |
| 4.2.2.1.1 - Simulação de Monte Carlo | 57 |
| 4.2.2.2 - Método de estimação de máxima verosimilhança | 68 |
| 4.2.2.2.1 - Simulação de Monte Carlo | 76 |
| 4.2.2.2.2 - Métodos numéricos | 85 |
| 4.2.2.2.2.1 - Método de Newton-Raphson | 85 |
| 4.2.2.2.2.2 - Algoritmo EM | 87 |
| 4.2.2.2.2.2.1 - Algoritmo EM com dados censurados à direita | 91 |
| 4.3 - Intervalo de confiança | 96 |
| 4.3.1 - Método da matriz de Fisher | 98 |
| 4.3.2 - Método da razão de verosimilhança | 99 |
| 4.3.3 - Método bootstrap | 101 |
| 4.3.3.1 - Método bootstrap não paramétrico | 102 |
| 4.3.3.2 - Método bootstrap paramétrico | 104 |
| 4.3.3.3 - Intervalo de confiança bootstrap | 104 |
| 4.3.3.3.1 - Intervalo de confiança bootstrap normal | 105 |
| 4.3.3.3.2 - Intervalo de confiança bootstrap-t | 106 |
| 4.3.3.3.3 - Intervalo de confiança bootstrap percentil | 107 |
| 4.3.3.3.4 - Intervalo de confiança bootstrap BCPB | 107 |
| 4.3.3.3.5 - Intervalo de confiança bootstrap BCa | 108 |
| 4.4 - Conclusões | 110 |
| Capítulo 5 - Aplicação experimental | 111 |
| 5.1 - Introdução | 111 |
| 5.2 - Bomba centrífuga | 111 |
| 5.3 - Caracterização do caso de estudo | 114 |
| 5.3.1 - Empanque mecânico | 117 |
| 5.3.1.1 - Estimação dos parâmetros da distribuição | 124 |
| 5.3.2 - Rolamentos | 131 |
| 5.3.2.1 - Estimação dos parâmetros da distribuição | 133 |

| | |
|--|------------|
| 5.4 - Conclusões | 139 |
| Capítulo 6 - Conclusões | 140 |
| 6.1 - Trabalhos futuros | 142 |
| Referências | 143 |
| Anexo A - Simulação de Monte Carlo | 152 |
| Anexo B - Valor esperado | 157 |
| Anexo C - Valores críticos de Kolmogorv-Smirnov | 159 |

Lista de figuras

| | |
|---|----|
| Figura 2.1 - Esquema de tipos de manutenção | 6 |
| Figura 2.2 - Ação de manutenção após surgir a falha (manutenção corretiva e curativa) | 6 |
| Figura 2.3 - Ação de manutenção preventiva condicionada, com inspeções periódicas | 8 |
| Figura 2.4 - Conceito da otimização da manutenção | 9 |
| Figura 2.5 - Metodologia de estimação paramétrica | 11 |
| Figura 2.6 - Classificação dos tipos de dados | 12 |
| Figura 2.7 - Tipos de dados censurados | 15 |
| Figura 3.1 - Relação entre $f(t)$, $F(t)$ e $R(t)$ | 25 |
| Figura 3.2 - Tempo entre falhas e tempo acumulado de funcionamento | 28 |
| Figura 3.3 - Representação de T_{in} e T_{out} | 30 |
| Figura 3.4 - Evolução da taxa de ocorrência com o tempo | 31 |
| Figura 3.5 - Possível tendência do tempo entre falhas | 35 |
| Figura 3.6 - Número acumulado de falhas em função do tempo t | 37 |
| Figura 3.7 - Modelo de processo de seleção para sistemas reparáveis | 42 |
| Figura 3.8 - O efeito de η na função densidade de probabilidade da distribuição de Weibull ($\beta = 3$) | 45 |
| Figura 3.9 - O efeito de β na função densidade de probabilidade da distribuição de Weibull ($\eta = 1$) | 46 |
| Figura 3.10 - O efeito de β na função intensidade de falha ($\eta = 1$) | 47 |
| Figura 4.1 - Comparação da estimação do parâmetro de forma, β , para dados completos ($\eta_T = \beta_T = 1$) | 60 |
| Figura 4.2 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = 1$, $\beta_T = 0.5$) | 61 |

| | |
|--|----|
| Figura 4.3 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = \beta_T = 1$) | 62 |
| Figura 4.4 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = 1, \beta_T = 3$) | 62 |
| Figura 4.5 - Comparação da estimação do parâmetro de forma, β , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 20\%$ | 64 |
| Figura 4.6 - Comparação da estimação do parâmetro de forma, β , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 60\%$ | 64 |
| Figura 4.7 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 20\%$ | 66 |
| Figura 4.8 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 60\%$ | 66 |
| Figura 4.9 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = 1, \beta_T = 3$), com $c = 20\%$ | 67 |
| Figura 4.10 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = 1, \beta_T = 3$), com $c = 60\%$ | 67 |
| Figura 4.11 - Comparação da estimação do parâmetro de forma, β , para dados completos ($\eta_T = \beta_T = 1$) | 78 |
| Figura 4.12 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = 1, \beta_T = 0.5$) | 79 |
| Figura 4.13 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = \beta_T = 1$) | 79 |
| Figura 4.14 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = 1, \beta_T = 3$) | 80 |
| Figura 4.15 - Comparação da estimação do parâmetro de forma, β , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 20\%$ | 81 |
| Figura 4.16 - Comparação da estimação do parâmetro de forma, β , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 60\%$ | 82 |
| Figura 4.17 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 20\%$ | 83 |
| Figura 4.18 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 60\%$ | 83 |
| Figura 4.19 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = 1, \beta_T = 3$), com $c = 20\%$ | 84 |
| Figura 4.20 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = 1, \beta_T = 3$), com $c = 60\%$ | 84 |
| Figura 4.21 - Procedimento do método de Newton-Raphson | 87 |
| Figura 4.22 - Procedimento do algoritmo EM | 89 |
| Figura 4.23 - Intervalo de confiança bilateral | 97 |

| | |
|---|-----|
| Figura 4.24 - Intervalo de confiança unilateral inferior e superior | 98 |
| Figura 4.25 - Intervalo de confiança pelo método da razão de verossimilhança | 100 |
| Figura 4.26 - Método bootstrap não paramétrico | 103 |
| Figura 4.27 - Método bootstrap paramétrico | 105 |
| Figura 5.1 - Classificação das bombas hidráulicas | 113 |
| Figura 5.2 - Árvore funcional da bomba centrífuga | 115 |
| Figura 5.3 - Gráfico de Pareto das bombas centrífugas por componente | 117 |
| Figura 5.4 - Empanque mecânico de uma bomba centrífuga | 118 |
| Figura 5.5 - Representação das forças envolvidas num empanque mecânico | 120 |
| Figura 5.6 - Representação da ação de extrusão no o-ring | 121 |
| Figura 5.7 - Gráfico de probabilidade de Weibull dos dados de falha em horas para um empanque mecânico | 125 |
| Figura 5.8 - Gráfico de probabilidade de Weibull dos dados de falha em dias para todos os empanques mecânicos | 126 |
| Figura 5.9 - Dispersão dos valores obtidos pelo método bootstrap para o empanque 1 - | 128 |
| Figura 5.10 - Densidade de probabilidade pelo método bootstrap para o empanque 1 - | 129 |
| Figura 5.11 - Principais causas de falha no rolamento de uma bomba centrífuga | 132 |
| Figura 5.12 - Gráfico de probabilidade de Weibull dos dados de falha em horas para um rolamento | 135 |
| Figura 5.13 - Gráfico de probabilidade de Weibull dos dados de falha em dias para todos os rolamentos | 136 |
| Figura 5.14 - Dispersão dos valores obtidos pelo método bootstrap para o rolamento 1 | 138 |

Lista de tabelas

| | |
|--|-----|
| Tabela 3.1 - Características típicas da função densidade de probabilidade e da função intensidade de falha em função da variação de β ----- | 48 |
| Tabela 4.1 - Valor estimado da função probabilidade acumulada pelo método de Kaplan-Meier ----- | 53 |
| Tabela 4.2 - Valor estimado da função probabilidade acumulada pelo método de Herd-Johnson ----- | 54 |
| Tabela 4.3 - Valor estimado da função probabilidade acumulada pelo método de Nelson ----- | 55 |
| Tabela 4.4 - Parâmetros da simulação para a estimação do parâmetro de forma, β , com dados completos ----- | 59 |
| Tabela 4.5 - Parâmetros da simulação para a estimação do parâmetro de escala, η , com dados completos ----- | 61 |
| Tabela 4.6 - Parâmetros da simulação para a estimação do parâmetro de forma, β , com dados censurados ----- | 63 |
| Tabela 4.7 - Parâmetros da simulação para a estimação do parâmetro de forma, η , com dados censurados ----- | 65 |
| Tabela 4.8 - Parâmetros da simulação para a estimação do parâmetro de forma, β , com dados completos ----- | 77 |
| Tabela 4.9 - Parâmetros da simulação para a estimação do parâmetro de escala, η , com dados completos ----- | 78 |
| Tabela 4.10 - Parâmetros da simulação para a estimação do parâmetro de forma, β , com dados censurados à direita tipo 1 ----- | 81 |
| Tabela 4.11 - Parâmetros da simulação para a estimação do parâmetro de forma, η , com dados censurados à direita tipo 1 ----- | 82 |
| Tabela 5.1 - Número e percentagem de falhas, por componente e comparação com (Bloch, 2010) ----- | 116 |
| Tabela 5.2 - Modos, causas e respectivos mecanismos de falha no empanque mecânico | 123 |
| Tabela 5.3 - Valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ (dias) para cada um dos empanques mecânicos pelo método dos mínimos quadrados ----- | 127 |

| | |
|--|-----|
| Tabela 5.4 - Valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ (dias) para cada um dos empanques mecânicos pelo método de máxima verosimilhança (EM) e respetivo intervalo de confiança pelo método bootstrap-t ----- | 128 |
| Tabela 5.5 - Modos, causas e respetivos mecanismos de falha nos rolamentos ----- | 134 |
| Tabela 5.6 - Valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ (dias) dos rolamentos para cada bomba centrífuga pelo método dos mínimos quadrados ----- | 136 |
| Tabela 5.7 - Valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ (dias) dos rolamentos para cada bomba centrífuga pelo método de máxima verosimilhança (EM) e respetivo intervalo de confiança pelo método bootstrap-t ----- | 137 |

Abreviaturas e Símbolos

Lista de abreviaturas (ordenadas por ordem alfabética)

| | |
|------|--|
| ABAO | <i>As bad as old</i> (condição de tão mal quanto antes) |
| AGAN | <i>As good as New</i> (condição de como novo) |
| B | Enviesamento |
| Cov | Covariância |
| EM | Algoritmo <i>Expectation-Maximization</i> |
| EQM | Erro quadrático médio |
| H-J | Estimador de Herd-Johnson |
| HPP | <i>Homogeneous Poisson process</i> (processo de Poisson homogéneo) |
| iid | Independente e identicamente distribuído |
| K-M | Estimador de Kaplan-Meier |
| KS | Teste Kolmogorov-Smirnov |
| MTBF | <i>Mean time between failures</i> (tempo médio entre falhas) |
| MTTF | <i>Mean time to failure</i> (tempo médio até à falha) |
| MTTR | <i>Mean time to repair</i> (tempo médio de reparação) |
| MV | Método de máxima verosimilhança |
| NHPP | <i>Non-homogeneous Poisson process</i> (processo de Poisson não homogéneo) |
| NR | Método de Newton-Raphson |
| RP | <i>Renewal process</i> (processo de renovação) |
| TTF | <i>Time to failure</i> (tempo até à falha) |
| Var | Variância |

Lista de símbolos

| | |
|------------|--|
| β | Parâmetro de forma (distribuição de Weibull) |
| γ | Parâmetro de localização (distribuição de Weibull) |
| Γ | Função Gama |
| Γ_d | Variável tempo (fim do registo, truncado) |
| Γ_e | Variável tempo (início do registo, truncado) |
| δ | Indicador de censura |

| | |
|---------------|---|
| η | Parâmetro de escala (distribuição de Weibull) |
| θ | Parâmetro da distribuição |
| $\lambda(t)$ | Função intensidade de falha |
| $v(t)$ | Taxa de ocorrência de falhas |
| $\Psi(t)$ | Função digamma |
| $\Psi'(t)$ | Função trigama |
| c | Porcentagem de dados censurados |
| C_d | Variável tempo (fim do registo) |
| C_e | Variável tempo (início do registo) |
| D | Disponibilidade |
| $f(t)$ | Função densidade de probabilidade |
| $F(t)$ | Função de probabilidade acumulada de falha |
| $h(t)$ | Função de risco |
| $H(t)$ | Função de risco acumulado |
| $L(\theta)$ | Função de verosimilhança |
| $\ln(\theta)$ | Função de log-verosimilhança |
| M | Número de iterações |
| n | Tamanho da amostra |
| r | Número de falhas |
| $R(t)$ | Fiabilidade |
| T | Variável tempo |

Capítulo 1

Introdução

As paragens devidas a falhas com os equipamentos são, geralmente, uma das principais causas de baixos níveis de desempenho e de baixa produtividade, o que conduz a que os custos das operações sejam agravados, retirando competitividade às organizações.

O conceito e a implementação da manutenção tem passado por profundas modificações, devido essencialmente a dois aspetos. Primeiro, a uma significativa evolução na sofisticação dos equipamentos utilizados nos processos produtivos, provocada principalmente pelas solicitações de aumento de produtividade e de qualidade como fator de competitividade e, segundo, às consequências de eventuais falhas, relativamente aos impactos económicos, ambientais e de segurança.

A necessidade de ter sistemas cada vez mais eficazes e seguros levou ao desenvolvimento, no contexto operacional, de estratégias de manutenção que conduzam à melhoria contínua da disponibilidade e da segurança operacional dos sistemas, com custos controlados

É neste contexto que se pretende evidenciar a importância do conhecimento do comportamento dos equipamentos. Esse conhecimento permite de forma fundamentada definir quais as políticas de manutenção mais adequadas a cada equipamento e aos componentes neles inseridos, de modo a reduzir significativamente os custos associados à manutenção pela otimização do número de inspeções, pela diminuição do número de componentes substituídos e de stock e pela melhor preparação dos trabalhos de manutenção e diminuição dos tempos de reparação, por exemplo.

As novas ferramentas computacionais de simulação são uma grande ajuda para o tratamento e análise de dados; no entanto, não bastam por si só o “software” ou o “hardware” para as análises de fiabilidade. É necessário saber com rigor e profundidade a forma como modelamos matematicamente a realidade das falhas de um equipamento ou componente. Ao longo desta tese, aprofundam-se as metodologias e os desenvolvimentos

matemáticos e estatísticos mais recentes, que permitem representar a realidade das falhas de um sistema, tendo em conta o seu histórico.

A gestão moderna utiliza cada vez mais as ferramentas de simulação para a tomada de decisões. No caso da fiabilidade é cada vez mais importante saber em que fase da vida útil o equipamento se encontra, afim de compreender se é necessária alguma mudança na atividade de manutenção ou se o equipamento necessita de ser substituído.

A simulação computacional permite também fazer as análises, estudar e comparar modelos de fiabilidade sem necessidade de recorrer a casos reais, assim deste modo, é possível compreender melhor os modelos escolhidos e desenvolvidos para o sistema em análise. A utilização da simulação computacional permite alcançar bons resultados em muito menos tempo e com um menor investimento.

Com o intuito de inferir acerca do comportamento futuro dos equipamentos e dos componentes, é possível efetuar-se um processo de predição de falhas que consiste, essencialmente, no ajustamento de modelos teóricos ao conjunto de dados. Deste modo, é possível saber atempadamente e com uma probabilidade elevada quando irá ocorrer uma falha.

Os dados do sistema em análise podem ser obtidos de várias possíveis fontes, nomeadamente de testes de laboratório ou do registo das ocorrências ao longo da sua utilização (registo histórico). O histórico de falhas é a melhor fonte de informação sobre o sistema em análise; contudo, os dados provenientes do registo histórico em contexto operacional apresentam informação com características específicas que dificultam a sua análise, nomeadamente, dados incompletos e reduzido número de dados.

O processo de análise e estimação é suportado por um conjunto de técnicas, métodos e procedimentos estatísticos que permitem modelar o sistema em análise através do ajustamento de uma distribuição, previamente definida, ou através do cálculo de uma função de distribuição própria.

A função de distribuição estatística retrata o comportamento de uma dada população e portanto é referente a um sistema específico.

Esta tese apresenta os principais métodos de estimação paramétrica, bem como as limitações que lhe são subjacentes perante condições reais, e que são perfeitamente representativos da larga maioria dos casos que ocorrem no contexto industrial, nomeadamente na presença de dados incompletos e com reduzido número de dados.

Assim, apresenta-se uma abordagem nova que permite responder às limitações apresentadas pelos métodos habitualmente utilizados.

1.1 - Estrutura da tese

O documento foi estruturado em seis capítulos, que seguem a sequência da metodologia de estimação paramétrica apresentada no trabalho e inicia-se com a introdução da tese apresentada no presente capítulo.

O capítulo 2 evidencia a importância da aplicação de métodos fiabilísticos, de modo a conhecer o comportamento do equipamento ou do componente em análise, como ferramenta de apoio à gestão da manutenção. Apresentam-se as vantagens de conjugar este procedimento nos diferentes tipos de manutenção.

Neste capítulo apresenta-se uma metodologia que permite a estimação dos parâmetros da distribuição fiabilística que melhor se ajusta aos dados obtidos. Na descrição dos passos da metodologia dá-se particular atenção aos principais problemas práticos que geralmente estão associados na análise de dados provenientes do registo histórico de falhas de equipamentos mecânicos, nomeadamente o reduzido número de dados e a existência de dados incompletos.

O capítulo 3 descreve com maior detalhe o passo 3 da metodologia de estimação paramétrica apresentada no capítulo 2, sendo referidos os diversos procedimentos para a seleção da distribuição que melhor se adequa para modelar um determinado conjunto de dados.

Neste capítulo, apresentam-se as noções de sistema reparável e de sistema não reparável, dando ênfase às suas diferenças e à distinta abordagem na sua aplicação nos estudos de fiabilidade.

No final do capítulo, caracteriza-se a distribuição estatística escolhida que melhor se adequa ao histórico de falhas do caso real em estudo, designadamente, a distribuição de Weibull.

O capítulo 4 apresenta os conceitos e fundamentos teóricos de alguns dos métodos de estimação dos parâmetros da distribuição de Weibull, divididos em dois grupos, métodos gráficos e métodos analíticos. Os métodos analíticos utilizados para estimar os parâmetros de Weibull são: o método dos mínimos quadrados e o método de máxima verosimilhança. É aplicada a simulação de Monte Carlo para comparar o desempenho dos vários métodos.

Como as equações de máxima verosimilhança em muitas situações não apresentam solução analítica, para determinar as suas soluções apresentam-se dois métodos numéricos de otimização para a sua resolução, designadamente o método de Newton-Raphson e o algoritmo Expectation-Maximization (EM).

A análise do algoritmo EM é feita para dados completos e incompletos. Contudo, é dada maior ênfase na presença de dados censurados à direita tipo 1, devido à sua aplicação para o caso em estudo.

No capítulo 5 mostra-se uma aplicação real da metodologia proposta. O sistema em estudo consiste em 5 bombas centrífugas de uma indústria petroquímica. O capítulo começa por descrever e aplicar um conjunto de procedimentos que permite identificar quais os componentes mais críticos do sistema e, assim, incidir especificamente sobre estes componentes a aplicabilidade da metodologia. Constata-se a existência de dados incompletos e de um número reduzido de dados.

A estimação dos parâmetros da distribuição é realizada pelo método dos mínimos quadrados e pelo método de máxima verosimilhança, através do algoritmo EM.

Para o método de máxima verosimilhança é apresentado o intervalo de confiança, dos parâmetros estimados, pelo método bootstrap.

Finalmente, no capítulo 6, são apresentadas as conclusões e apontadas algumas perspectivas de trabalhos futuros, com base no trabalho desenvolvido.

Capítulo 2

Metodologia de estimação paramétrica

2.1 - Introdução

O capítulo apresenta os objetivos da tese, que consiste na estimação dos parâmetros da distribuição fiabilística que melhor se ajusta aos dados provenientes de um registo histórico, de modo a conhecer o comportamento do equipamento ou do componente analisado. A aplicação num caso prático real, de onde advêm as condicionantes críticas que na maioria dos casos reais acontecem, designadamente, o reduzido número de dados e dados incompletos, é feita de modo a exemplificar a metodologia apresentada.

Inicialmente, apresenta-se as vantagens de conjugar este procedimento nos diferentes tipos de manutenção. Posteriormente, descrevem-se as várias etapas da metodologia para a estimação dos parâmetros da distribuição.

Realça-se a importância do registo histórico dos sistemas e das características específicas que dificultam a análise dos dados recolhidos.

2.2 - Enquadramento

A manutenção é uma combinação de vários fatores de acordo com os objetivos pretendidos.

Existem, essencialmente, três tipos de manutenção: manutenção corretiva/ curativa, preventiva sistemática e preventiva condicionada. Pode-se generalizar e descrever os diferentes tipos de manutenção segundo a forma ilustrada na figura 2.1.

Se o tipo de manutenção implementada for essencialmente corretiva/ curativa, em que a ação de manutenção intervem após a ocorrência de uma falha, com o objetivo de remediar a situação (curativa) ou repor as condições nominais (corretiva), verificam-se custos elevados associados à paragem de produção, em função da intervenção, para restituir o funcionamento

do equipamento.

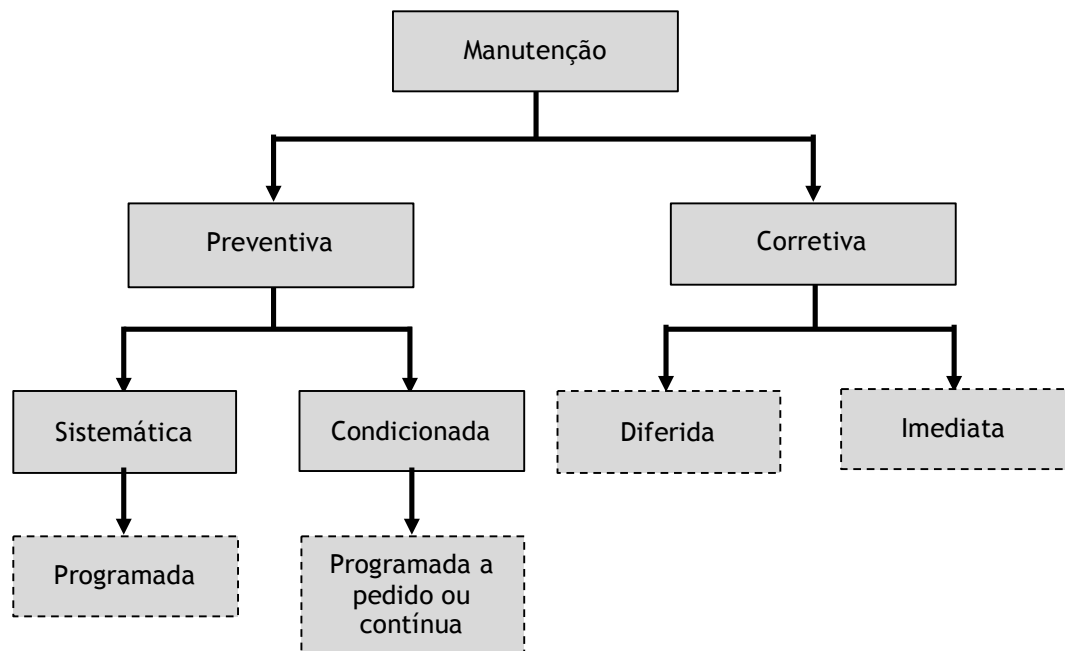


Figura 2.1 - Esquema de tipos de manutenção (NP EN 13306, 2007).

O custo de perda de produção depende da duração do intervalo de tempo durante o qual o serviço deixa de ser assegurado por falta de equipamentos disponíveis e da importância do equipamento no ciclo produtivo, conforme referido na figura 2.2.

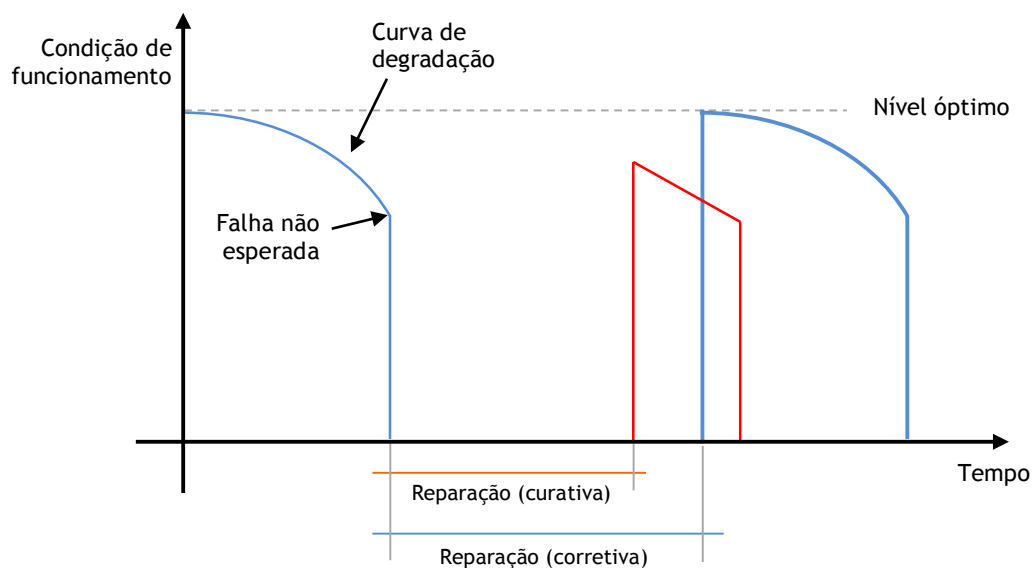


Figura 2.2 - Ação de manutenção após surgir a falha (manutenção corretiva e curativa) (Monchy, 2003).

Podem ainda estar agregados graves consequências da exposição de pessoas e do meio ambiente decorrente das falhas, por exemplo, altas temperaturas, vazamentos e contaminação.

Para diminuir os efeitos na produção, criados pelas falhas inesperadas dos equipamentos, uma empresa que use exclusivamente a manutenção corretiva e curativa deve ter capacidade para reagir imediatamente a todas as falhas. Assim, é necessário manter em stock componentes ou equipamentos de reserva, pelo menos os mais importantes para os processos mais críticos.

Para diminuir os custos associados com a manutenção corretiva/ curativa e também para diminuir o impacto da ocorrência no sistema produtivo, é possível planejar previamente as intervenções mais importantes ou as que têm um carácter repetitivo, estipulando os meios materiais e humanos necessários, e programar as tarefas respetivas.

A manutenção preventiva sistemática é uma das metodologias que se enquadram nesta perspectiva, sendo uma abordagem desenvolvida para evitar as falhas inesperadas. É planeada para preservar e melhorar a fiabilidade dos equipamentos pela substituição de componentes sujeitos a desgaste antes que estes falhem.

Pelo facto de as intervenções serem periódicas, este tipo de manutenção implica uma matriz organizativa forte e um planeamento eficiente.

Na manutenção preventiva sistemática é possível intervir de duas formas:

- Substituição dos componentes em intervalos regulares (manutenção sistemática);
- Substituição dos componentes quando o equipamento não está a funcionar por outros motivos (manutenção de oportunidade).

A implementação desta prática acaba por reduzir as consequências das falhas mas eleva o custo financeiro com a substituição de componentes que poderiam ainda estar em bom estado, além de incorrer na possibilidade de introduzir falhas no equipamento devido à intervenção da manutenção.

As ações de manutenção preventiva condicionada iniciam-se quando se deteta o princípio do processo de degradação do equipamento, através do controlo de parâmetros que estão associados à condição do seu funcionamento e refletem a sua degradação.

Neste tipo de manutenção, deve-se proceder ao acompanhamento das condições de funcionamento dos equipamentos através da inspeção dos parâmetros que o caracterizam, de modo a detetar as situações em que se ultrapassam os valores de referência para os parâmetros seleccionados.

A realização de inspeções periódicas, conforme representado na figura 2.3 ou a monitorização contínua das condições operacionais dos equipamentos, permite evitar a ocorrência de falhas.

A taxa de falhas do equipamento diminui, como resultado da substituição de componentes em função da verificação do seu estado de funcionamento.

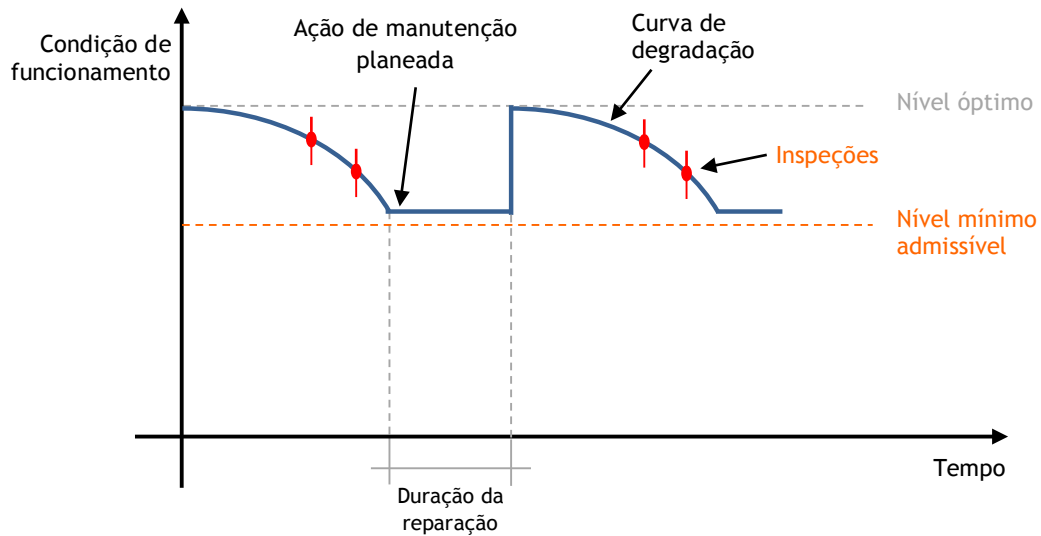


Figura 2.3 - Ação de manutenção preventiva condicionada, com inspeções periódicas (Monchy, 2003).

O acompanhamento do desempenho dos equipamentos é realizado através da monitorização de grandezas que estão envolvidas no sistema em análise, por exemplo, vibrações, temperatura e corrente elétrica. O mecanismo ideal é fazer a aquisição da informação em regime contínuo e a avaliação automaticamente, estabelecendo níveis de alarme para as variáveis analisadas. A monitorização deve fornecer dados suficientes para uma análise de tendência e assim permitir a emissão de um diagnóstico para a tomada de decisão. Contudo, fazer a aquisição em regime contínuo de todos os equipamentos não é economicamente viável. Assim, este tipo de acompanhamento é limitado aos equipamentos mais críticos.

O aumento da frequência de atividades de inspeção do estado dos equipamentos diminui a exposição ao risco e as consequências das falhas, contudo, aumenta o custo direto da manutenção.

A implementação de uma política de manutenção requer a otimização dos três tipos de manutenção (manutenção corretiva e curativa, manutenção preventiva sistemática e manutenção preventiva condicionada), de forma a que sejam aplicados apenas quando e onde necessários, eliminando toda a manutenção excessiva.

O tipo de manutenção de um componente ou equipamento depende também da severidade das consequências provocadas pela sua falha. Para os componentes cujas consequências das falhas são pouco significativas é preferível, em alguns casos, apenas submetê-los a ações de manutenção corretiva quando a falha ocorre. Na situação oposta, os componentes ou equipamentos cujas falhas podem resultar em situações críticas de segurança ou económicas, a manutenção preventiva sistemática ou condicionada deve ser aplicada de forma a evitar a ocorrência de falhas.

A prática da manutenção preventiva diminui as paragens inesperadas e aumenta o controlo sobre os equipamentos. Contudo, se as ações de manutenção forem excessivas, o custo resultante será elevado. Assim, quando se pretende atingir o máximo desempenho dos equipamentos, todos os tipos de manutenção devem ser considerados e os custos envolvidos devem ser ponderados (Jardine, 2013).

Um dos objectivos principais da otimização da manutenção é determinar qual a estratégia de manutenção que minimiza os custos. Esta estratégia deve fornecer o melhor equilíbrio possível entre os custos diretos de manutenção e as consequências ou penalizações de não realizar a manutenção conforme necessário, sem prejuízo para a saúde, segurança e meio ambiente. O conceito da otimização da manutenção é ilustrado na figura 2.4.

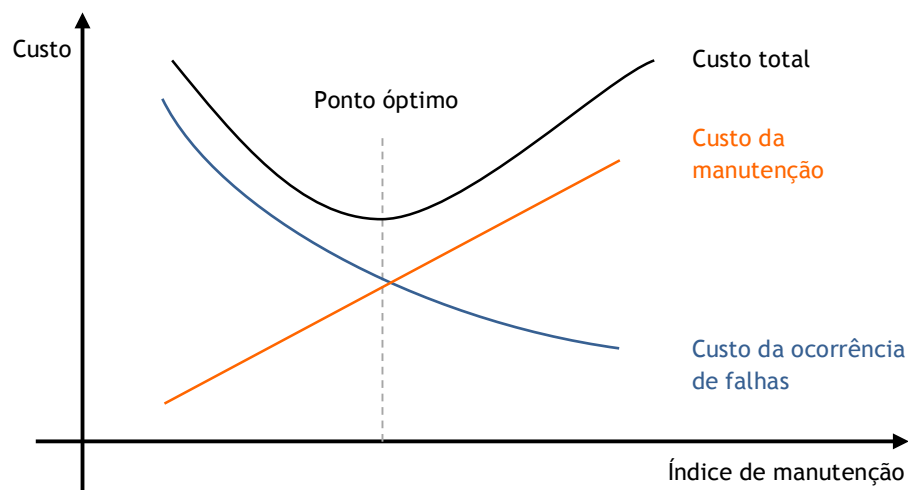


Figura 2.4 - Conceito da otimização da manutenção (Ferreira, 1998).

A otimização visa a combinação destes fatores com o objetivo de determinar o nível ideal da manutenção. Este é geralmente obtido no ponto inferior da curva correspondente ao custo total do impacto da manutenção (custo direto da manutenção mais o custo das consequências da ocorrência de falhas) conforme indicado na figura 2.4.

Em atividades em que o custo não é o fator prioritário em detrimento, por exemplo, da segurança ou da qualidade do produto a abordagem será diferente.

Os custos diretos da manutenção incluem essencialmente os custos da mão-de-obra, dos equipamentos e das ferramentas para a inspeção e a reparação, do consumo de matéria-prima para a manutenção (consumíveis), dos componentes ou equipamentos a substituir, dos contratos de manutenção e dos trabalhos subcontratados (Ferreira, 1998).

Os custos das consequências das falhas incluem os custos de perda de produção, os custos associados aos prazos não cumpridos ou a degradação da imagem no mercado, o custo dos colaboradores sem ocupação e de amortização dos equipamentos parados. Em certos casos, acrescenta-se ainda, o custo associado à eventual danificação ou destruição, no decorrer da falha, de outros itens inseridos no sistema e também o custo de perda de qualidade do produto ou serviço antes de ocorrer a avaria (Ferreira, 1998).

Nesta perspetiva, qualquer que seja o tipo de manutenção utilizada, é possível utilizar uma ferramenta que se torna essencial no eficiente desempenho da ação de manutenção, através do conhecimento do comportamento do equipamento ou do componente pela utilização de métodos fiabilísticos com base no registo histórico de falhas.

Este é o objetivo central deste trabalho, que consiste na caracterização da distribuição fiabilística que melhor se ajusta aos dados existentes para que seja possível estimar a ocorrência das falhas dentro de um intervalo temporal de confiança que é quantificável e aplicar o método a um caso prático real com características semelhantes ao que ocorre na maioria dos casos reais, em que o registo histórico de falhas é constituído por poucos dados e com dados incompletos, conforme apresentado no capítulo 5.

O conhecimento do comportamento do equipamento ou do componente permite reduzir significativamente os custos associados à manutenção pela otimização do número de inspeções, da diminuição do número de componentes substituídos e de stock, melhor preparação dos trabalhos de manutenção, diminuição dos tempos de reparação, por exemplo.

O processo de estimação é um conjunto de técnicas, métodos e procedimentos estatísticos para analisar os dados sobre a variável de interesse, que pode ser o tempo que decorre desde o instante inicial bem definido, como por exemplo, a instalação de um equipamento, até à ocorrência de um determinado acontecimento, nomeadamente uma falha do equipamento ou do componente em estudo.

Deste modo, é necessário identificar qual o modelo teórico que melhor se ajusta e representa os dados obtidos. A análise paramétrica e a não paramétrica são dois métodos possíveis para implementar o modelo teórico do sistema.

A análise paramétrica pressupõe que os dados obtidos se ajustam a uma distribuição específica como, por exemplo, a distribuição de Weibull. A análise não paramétrica não requer o conhecimento da distribuição do tempo até à falha mas é necessário determinar as funções empíricas da fiabilidade a partir dos dados. A análise não paramétrica é, geralmente, menos eficiente quando prevalecem as condições nas quais a análise paramétrica é válida (Guimarães, 1999).

Quando a análise paramétrica é a escolhida, o estudo visa estimar os parâmetros da distribuição que permitem modelar o sistema em análise. Os parâmetros são características da distribuição que retrata o comportamento de uma dada população e portanto fixos para um sistema específico. Assim, a estimativa dos parâmetros do sistema é obtida a partir dos dados recolhidos da população e para a sua determinação emprega-se uma formulação, a qual é denominada de estimação.

Nesta tese foi escolhida a análise paramétrica pelo motivo referido anteriormente, ou seja, quando prevalecem as condições nas quais a análise paramétrica é válida, o que acontece no caso analisado no capítulo 5, a análise paramétrica é escolhida porque é mais eficiente.

O capítulo seguinte ilustra o procedimento adotado para a estimação dos parâmetros da distribuição que melhor se ajusta aos dados existentes.

2.3 - Metodologia de estimação paramétrica

A metodologia de estimação paramétrica pode ser descrita de acordo com as cinco etapas representadas na figura 2.5.

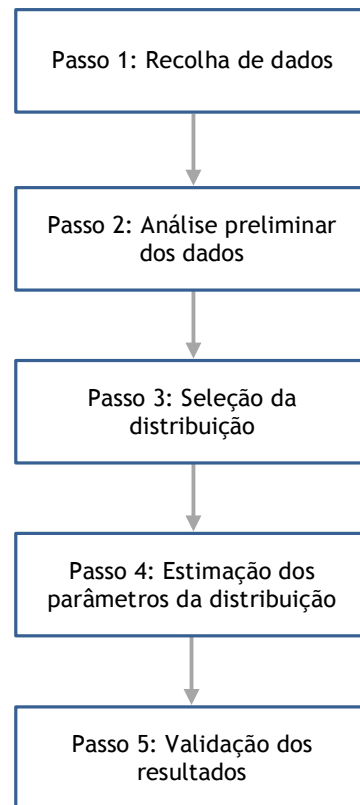


Figura 2.5 - Metodologia de estimação paramétrica.

2.3.1 - Passo 1: Recolha de dados

Os dados do sistema em análise podem ser obtidos de diversas fontes, nomeadamente, de testes de laboratório ou do registo das ocorrências ao longo da sua utilização (registo histórico).

Neste estudo os dados são provenientes do registo de falhas ao longo da utilização do equipamento em análise.

É importante referir que um adequado registo histórico de falhas é fundamental para o acompanhamento de um equipamento, bem como, ao correto desempenho das atividades referentes à sua manutenção. Para que os dados obtidos possam proporcionar indicadores válidos para a tomada de decisões é necessário que a sua recolha seja feita de forma criteriosa e homogénea. Só assim, é possível obter resultados precisos e consistentes relativamente aos eventos ocorridos no sistema.

O histórico de falhas é a melhor fonte de informação sobre o sistema em análise. Quando esta informação não existe ou é inadequada, é possível utilizar a informação existente em

bancos de dados externos. Esta informação é editada por diversos organismos e indicada para os componentes mais comuns, geralmente, o MTTF e a intensidade de falha. Porém, a utilização desta informação deve ser utilizada com algum cuidado, dado que a obtenção dos dados baseia-se em condições específicas de funcionamento que nem sempre se verificam.

Uma das mais importantes fontes de dados de fiabilidade é a base de dados da OREDA (Offshore Reliability Database). Este grupo de trabalho desenvolveu uma base de dados com informações proveniente de ações de manutenção de equipamentos de exploração e produção, a partir de uma ampla variedade de áreas geográficas, instalações, tipos de equipamentos e condições de funcionamento. Os dados foram obtidos principalmente em equipamentos localizados em plataformas industriais offshore do Mar do Norte e do Mar Mediterrâneo (Oreda, 2002).

2.3.2 - Passo 2: Análise preliminar dos dados

Para escolher a distribuição mais indicada para representar os dados obtidos é necessário fazer uma análise prévia da informação, isto é, é necessário caracterizar os dados registados cuja informação pode ser classificada conforme indicado na figura 2.6 (Biolini, 2014), (Ebeling, 1997).

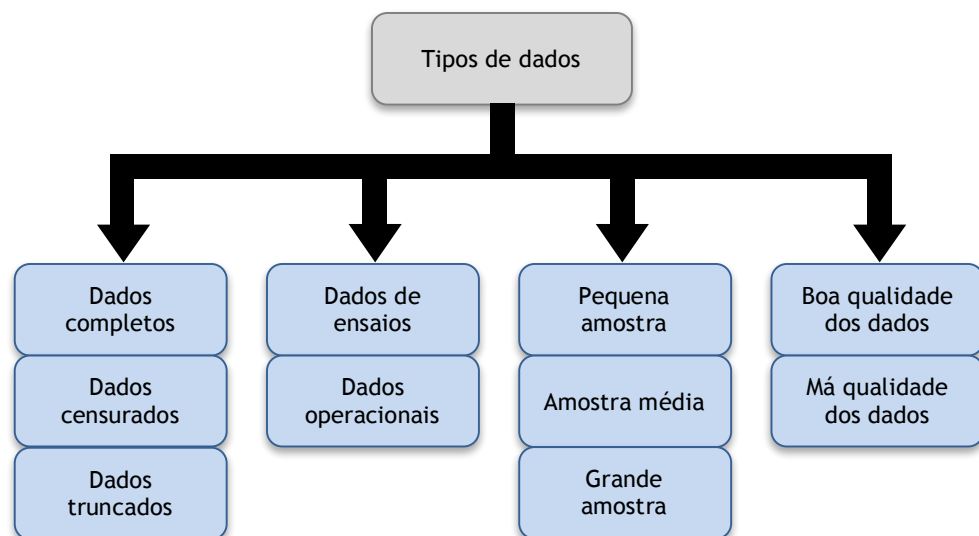


Figura 2.6 - Classificação dos tipos de dados (Adaptado de Biolini, 2014 e Ebeling, 1997).

Relativamente à informação obtida através de ensaios ou no contexto operacional, deve-se considerar que dificilmente se obtêm todos os elementos de um sistema e, portanto, devem-se obter estimativas dos parâmetros da distribuição em função dos dados recolhidos.

Se todas as unidades são analisadas até à falha, então os dados obtidos são considerados completos, caso contrário os dados são considerados incompletos ou parciais (Lawless, 2003).

Os dados são considerados completos quando é conhecido o tempo exato de cada falha do sistema. Em muitos casos, os dados obtidos contêm incertezas, ou seja, não é conhecido o momento exato em que ocorreu a falha. Os dados que contêm essa incerteza relativamente ao momento em que ocorreu o evento, são considerados incompletos ou parciais. Os dados incompletos podem ser classificados em censurados ou truncados.

Os dados incompletos dão somente parte da informação sobre o tempo de falha das unidades em análise. Contudo essa informação não deve ser ignorada ou tratada como falha. Sem a presença desses dados, não seria possível realizar uma boa estimação de parâmetros e assim fazer uma análise adequada.

Os dados incompletos podem ser classificados conforme indicado na figura 2.7 (Gijbels, 2010), (Lawless, 2003), (Tobias, 2011).

2.3.2.1 - Dados censurados à direita

Seja T uma variável aleatória que representa o tempo de falha e C_d uma outra variável aleatória independente de T que corresponde à conclusão do registo da informação. Diz-se que o tempo de falha t_i é censurado à direita quando não se conhece o seu valor exato apenas se sabe que é superior a c_d . Portanto,

$$t_i = \min(T, C_d) \text{ e } \delta_i = \begin{cases} 1 & \text{se } t_i \leq c_d \\ 0 & \text{se } t_i > c_d \end{cases}$$

A variável δ_i (indicador de censura) indica se t_i é censurado ou não. Se $t_i = c_d$ a informação é censurada à direita.

Os dados obtidos podem ser representados pelo par (t_i, δ_i) sendo t_i o tempo de falha ou de censura e δ_i a variável que indica se é referente a uma falha ou censura, isto é,

$$\delta_i = \begin{cases} 1, & \text{para dados não censurados} \\ 0, & \text{para dados censurados à direita} \end{cases}$$

Nos dados censurados à direita o tempo de falha das unidades com dados censurados apenas se sabe que é superior ao tempo de funcionamento correspondente à conclusão do registo da informação. Estes dados censurados à direita são ainda classificados em censura Tipo 1 se o registo da informação é interrompido a um tempo pré-determinado e censura Tipo 2 se o registo é concluído quando ocorrer um pré-determinado número de falhas.

2.3.2.1.1 - Dados censurados à direita Tipo 1

Na censura do Tipo 1 todas as unidades de um sistema são observadas até ao momento de conclusão do estudo. Para este esquema de censura o tempo em que cada unidade fica sob observação é fixo, enquanto que o número de unidades que falham (observações não censuradas) é aleatório. Este esquema de censura é utilizado quando se programa o tempo de duração do estudo.

2.3.2.1.2 - Dados censurados à direita Tipo 2

Na censura do Tipo 2 todas as unidades de um sistema são observadas até à ocorrência de um determinado número de falhas. Para este esquema de censura o número de falhas do sistema é fixo, enquanto que o tempo de observação é aleatório. Este esquema é utilizado quando se tem pouca ou nenhuma informação sobre a durabilidade da unidade.

2.3.2.2 - Dados censurados à esquerda

Analogamente, seja T uma variável aleatória que representa o tempo de falha e C_e uma outra variável aleatória independente de T que corresponde ao início do registo da informação. Diz-se que o tempo de falha t_i é censurado à esquerda quando não se conhece o seu valor exato apenas se sabe que é anterior a c_e , relativamente à falha i ($i = 1, 2, \dots, n$). Portanto,

$$t_i = \max(T, C_e) \text{ e } \delta_i = \begin{cases} 1 & \text{se } t_i \geq c_e \\ 0 & \text{se } t_i < c_e \end{cases}$$

Se $t_i = c_e$ a informação é censurada à esquerda.

Ou seja, os dados são considerados censurados à esquerda se o tempo de falha é anterior ao tempo do início do registo. A falha já aconteceu quando a unidade foi observada. Se a falha da unidade surge antes do início do estudo, o tempo de falha é somente conhecido após certo tempo. A falha ocorreu algum tempo antes do registo, mas não há a informação de exatamente quando.

2.3.2.3 - Dados censurados por intervalo

A censura por intervalo pode ser considerada como um caso mais geral de censura. Acontece quando por algum motivo não foi possível observar o tempo exato da falha, mas sim a ocorrência num certo intervalo de tempo. Este tipo de dados pode surgir quando o estado da unidade é verificada a cada inspeção. Se ocorreu uma falha entre inspeções não é

conhecido o exato momento em que a falha ocorreu, mas apenas que a falha ocorreu dentro do intervalo de tempo entre inspeções.

Os tipos de dados mencionados anteriormente são ilustrados na figura 2.7.

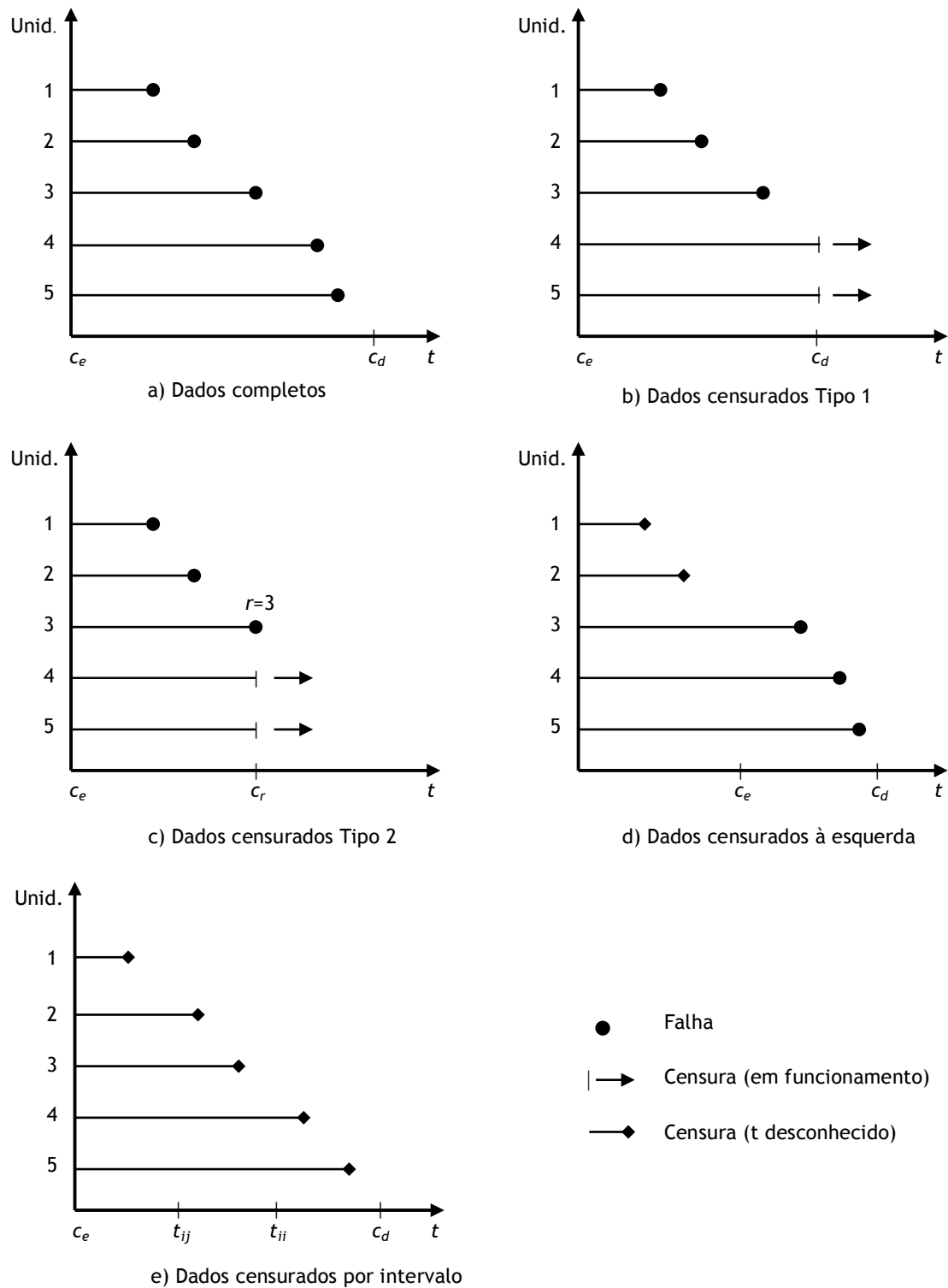


Figura 2.7 - Tipos de dados censurados (c_e - início do estudo, c_d - fim do estudo, c_r - t para $r=3$ e t_{ij} - inspeções).

Contudo existem outras possíveis classificações, nomeadamente, censura múltipla, censura progressiva, censura dupla e censura aleatória.

2.3.2.4 - Dados truncados

Por vezes, devido a restrições práticas ou para simplificar o modelo utilizado apenas são considerados os eventos que ocorrem dentro de uma janela temporal de observação. Todos os eventos que ocorrem fora deste espaço de tempo não são observados e consequentemente a sua existência não é conhecida. Os dados que têm este enquadramento são considerados truncados.

Os dados truncados podem ser classificados em dados truncados à direita e truncados à esquerda.

Os dados são considerados truncados à esquerda quando ocorrem antes do início do intervalo de observação e não são considerados para o estudo do sistema. Consequentemente se T é uma variável aleatória que representa o tempo de falha e T_e é uma variável aleatória independente de T que corresponde ao início do registo da informação, apenas são considerados para o estudo se $T \geq T_e$.

Analogamente, os dados são considerados truncados à direita quando ocorrem depois do início do intervalo de observação e não são considerados para o estudo do sistema. Consequentemente se T_d é uma variável aleatória independente de T que corresponde ao fim do registo da informação, apenas são considerados para o estudo se $T \leq T_d$.

Para além da classificação referida anteriormente, que divide os dados em completos e incompletos (censurados e truncados), os dados também podem ser classificados em diferentes grupos com base na origem dos dados, no tamanho da amostra e na qualidade dos dados, conforme referido na figura 2.6.

Do ponto de vista da origem, os dados podem ser classificados em dados obtidos de ensaios ou no contexto operacional (em fase de exploração). Com base no número de observações ou no tamanho da amostra, um conjunto de dados pode ser classificado em pequena, média ou grande dimensão.

Normalmente, um conjunto com menos de 20 dados é considerado como um conjunto de pequena dimensão (Abernethy, 2006).

Os dados podem também ser divididos em dados de boa ou má qualidade. Dados de boa qualidade, idealmente, não têm erros de medição nas observações ou o erro é pequeno o suficiente para ser desprezado, enquanto os dados de má qualidade envolvem valores atípicos ou observações em falta por exemplo.

Para certos equipamentos, a falha é catastrófica, sendo claro o seu ponto e modo de ocorrência. Noutros casos, a degradação lenta da capacidade funcional impede a clareza quanto à ocorrência do fim de vida e respetiva qualidade de informação. Outro exemplo são os dados obtidos em exploração, muitas vezes, devido aos métodos de recolha e registo são

agrupados por intervalos. Os tempos específicos de falha não são conhecidos devidamente o que implica perda na qualidade de informação.

Verifica-se que os dados obtidos pelo registo histórico do sistema ou pelos testes de laboratório apresentam características específicas que dificultam a sua análise.

Conforme referido anteriormente, a presença de dados censurados é muito comum. Porém outros problemas práticos estão geralmente associados na recolha dessa informação, nomeadamente,

a) Dados escassos:

Um dos principais problemas associados com a análise de dados é a falta de informação suficiente para executar corretamente as análises estatísticas. De facto, verifica-se que as metodologias estatísticas são limitadas quando realizadas com base num número reduzido de dados.

Os dados provenientes dos tempos de falhas dos equipamentos são normalmente reduzidos, o que enfatiza a necessidade de desenvolver métodos para lidar adequadamente com pequenos conjuntos de dados. Naturalmente, quanto maior for o conjunto de dados, mais precisa será a análise estatística. Por isso, o conteúdo da informação obtida pode ser incorreto sem uma análise cuidada.

Poder-se-á afirmar, que este problema referente à quantidade insuficiente de dados relativos aos tempos de falha dos equipamentos será uma constante, dado que o objectivo da manutenção é fazer com que as falhas sejam eventos raros. Deve-se esperar que com o aumento do desempenho da manutenção menos falhas devam ocorrer.

b) Efeito das ações de manutenção:

Normalmente, na prática, os componentes podem ser reparados ou ajustados, ao invés de substituídos, sempre que ocorre uma falha. Essas intervenções são suscetíveis de modificar os dados de fiabilidade do componente. Por isso, pode-se argumentar que o tempo de falha esperado após uma intervenção de reparação é diferente do tempo previsto para a primeira falha de um novo componente.

A abordagem fiabilística utilizada para o tempo de falha de um componente não reparável é equivalente para o tempo da primeira falha de um sistema reparável.

c) Qualidade dos dados:

Outro problema que está associado com a recolha dos dados é a qualidade da informação registada. A má qualidade da informação registada é devida a vários fatores, por exemplo, a dificuldade que pode surgir em identificar o modo de falha, a falta de formação do operador responsável pelo preenchimento do registo, a

informação incorreta do tempo de funcionamento do sistema em análise e a conservação do arquivo dos registros.

d) Combinação de dados:

Uma alternativa válida quando os dados são escassos é a combinação ou agrupamento de dados de componentes de outros equipamentos idênticos.

Este é um procedimento normalmente utilizado na análise da fiabilidade, especialmente em operações em que um grande número de equipamentos idênticos são utilizados.

A palavra "idêntico" pode ser substituída por "semelhante", em muitos casos. Deve-se ter em atenção a avaliação das semelhanças entre dois ou mais equipamentos antes de serem analisados em conjunto. Outro fator a ter em atenção é a semelhança das condições de funcionamento. Equipamentos idênticos, mas com condições de funcionamento diferentes, podem apresentar resultados completamente díspares.

Se for necessária uma análise mais rigorosa, quando na presença de duas ou mais amostras de dados a partir de diferentes equipamentos, vários métodos estatísticos podem ser utilizados para determinar se há diferenças significativas entre as amostras, mesmo em conjuntos de dados fortemente censurados.

Normalmente, os dados obtidos em ensaios têm algumas características especiais. Por exemplo, pequenos conjuntos de dados e dados censurados são muito comuns devido à necessidade de reduzir o tempo e o custo associado. O aumento do número de sistemas altamente fiáveis também conduz à dificuldade de recolha dos dados de falha. Um modo de contornar este obstáculo é a utilização de testes de vida acelerados. São então usadas relações matemáticas (propostas ou existentes) para extrapolar os resultados obtidos nos testes de vida acelerados para as condições ambientais usuais.

Os dados obtidos no contexto operacional, são maioritariamente caracterizados por pequenos conjuntos de dados, com combinações de diferentes tipos de censura e com má qualidade.

2.3.3 - Passo 3: Seleção da distribuição

A seleção da distribuição envolve a escolha do modelo adequado para modelar um determinado conjunto de dados. Perante a análise prévia das características dos dados realizada no passo 2, existem diversos meios através dos quais será possível antever qual a distribuição que presumivelmente melhor se ajusta.

Os procedimentos distinguem entre sistemas reparáveis e não reparáveis e permite avaliar os parâmetros estatísticos que evidenciem processos com taxa de falha crescente, decrescente ou constante e assim delimitar os modelos que teoricamente melhor se adequam a cada fenómeno, conforme apresentado no capítulo 3.

Para dados completos através dos parâmetros de localização e de assimetria é também possível determinar qual a melhor distribuição, por exemplo, se a média e a mediana forem iguais a distribuição é simétrica, logo a opção recai para uma distribuição normal ou de Weibull com $\beta = 3,5$.

A fonte dos dados, muitas vezes, também fornece uma pista para a seleção da distribuição apropriada, por exemplo, a distribuição lognormal ou a distribuição de Weibull têm sido utilizadas para modelar as falhas devido à fadiga e a distribuição exponencial para a falha de componentes eletrônicos.

2.3.4 - Passo 4: Estimação dos parâmetros da distribuição

Uma vez selecionada a distribuição é necessário estimar os parâmetros do modelo.

Têm sido desenvolvidas uma grande variedade de técnicas, das quais se podem destacar duas famílias, os métodos gráficos e analíticos. Podem ainda ser consideradas outras técnicas igualmente importantes, como por exemplo, os métodos Bayesianos.

A precisão da estimativa é dependente do tamanho dos dados e do método utilizado.

No capítulo 4 é apresentado o método de estimação considerado mais adequado ao caso prático analisado e é comparado com outro método habitualmente utilizado.

O estimador permite calcular uma estimativa de um determinado parâmetro baseado em dados observados.

Se o parâmetro é representado por θ então o estimador é normalmente escrito pela adição de um circunflexo sobre o símbolo, $\hat{\theta}$.

Diferentes métodos de estimação de parâmetros podem gerar estimativas amplamente diferentes, por isso, é importante ter critérios objetivos para a seleção de um método em detrimento das outras alternativas. O desempenho de cada estimador pode ser verificado pelas suas propriedades, por exemplo, enviesamento, variância e consistência.

Os seis critérios ou propriedades mais importantes para a seleção e comparação dos métodos de estimação são descritos a seguir (Guimarães, 1999), (Montgomery, 2003), (Rohde, 2014), (Tobias, 2011).

a) Enviesamento:

O enviesamento de $\hat{\theta}$ é definido como,

$$\delta(\hat{\theta}) = E(\hat{\theta}) - \theta \quad 2.1$$

Representa a diferença entre o valor esperado e o valor do parâmetro que se estima. Um estimador é considerado não enviesado se o seu valor esperado coincide com o valor real, ou seja,

$$E(\hat{\theta}) = \theta \Rightarrow \delta(\hat{\theta}) = 0$$

Um valor relativamente alto do enviesamento significa que o valor médio dos resultados está afastado do valor real. Um valor relativamente baixo significa que o valor médio está próximo do valor real.

b) Variância:

A variância de $\hat{\theta}$ é o valor esperado do quadrado da diferença entre os valores das estimativas e a média deles, ou seja:

$$\text{var}(\hat{\theta}) = E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2\right] \quad 2.2$$

Este atributo é usado para indicar o quanto distante, em média, o conjunto de estimativas está do valor esperado das estimativas. Se o valor da variância for relativamente alto significa que os valores estão dispersos. Se a variância for relativamente baixa significa que os valores estão agrupados.

Mesmo que a variância seja baixa o resultado pode estar longe do valor real. Se a variância for alta, o conjunto disperso de resultados pode estar próximo do valor real. Se todos os resultados estiverem afastados do valor correto, mas tiverem o mesmo valor a variância é zero.

c) Erro quadrático médio:

O erro quadrático médio de $\hat{\theta}$ é definido como o valor esperado do quadrado da diferença entre o valor da estimativa e o valor do parâmetro.

$$EQM(\hat{\theta}) = E\left[\left(\hat{\theta}(x) - \theta\right)^2\right] \quad 2.3$$

Esta propriedade permite indicar a que distância, em média, o conjunto de estimativas está do parâmetro a ser estimado. Um valor relativamente alto do EQM significa que a distância média dos resultados ao valor real é alta. Um valor relativamente baixo do EQM significa que a distância média dos resultados ao valor real é baixa.

O enviesamento, a variância e o EQM estão relacionados:

$$EQM(\hat{\theta}) = \text{var}(\hat{\theta}) + \left(\delta(\hat{\theta})\right)^2 \quad 2.4$$

Em particular, para um estimador não enviesado, o EQM é igual à variância.

d) Consistência:

Um estimador é consistente se, à medida que o tamanho da amostra aumenta, o valor esperado converge para o valor real e a variância converge para zero

$$\text{i) } \lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$$

$$\text{ii) } \lim_{n \rightarrow \infty} \text{var}(\hat{\theta}) = 0$$

e) Eficiência:

Dados dois estimadores $\hat{\theta}_1$ e $\hat{\theta}_2$ para um parâmetro θ , $\hat{\theta}_1$ é mais eficiente do que $\hat{\theta}_2$ se:

$$\text{var}(\hat{\theta}_1) < \text{var}(\hat{\theta}_2)$$

f) Simplicidade:

O método não envolve cálculos complicados e conhecimentos estatísticos sofisticados. Em suma, o método pode ser facilmente compreendido e de fácil aplicação.

Conclui-se que a situação ideal é ter um estimador não enviesado com baixa variância.

Muitas vezes, se apenas um pequeno enviesamento é permitido, então um estimador pode ser encontrado com o EQM baixo e com poucas estimativas da amostra discrepantes.

2.3.5 - Passo 5: Validação dos resultados

Após a identificação da distribuição teórica que supostamente melhor representa os dados e a determinação dos respetivos parâmetros, o passo seguinte e último do processo de estimação paramétrica é validar os resultados. Este passo consiste em efetuar a verificação da qualidade do ajuste da distribuição encontrada com os dados considerados.

Existem vários métodos para fazer este passo nomeadamente o teste Qui Quadrado, o teste Kolmogorov-Smirnov, o teste Anderson-Darling e o teste Cramer-Von Mises (O'Connor, 2012), (Kececioglu, 2002).

O teste utilizado neste estudo foi o teste de Kolmogorov-Smirnov (K-S) porque é mais simples de aplicar e apresenta resultados mais fiáveis para amostras com quantidade reduzida de dados (O'Connor, 2012), (Kececioglu, 2002).

No capítulo 5 o teste Kolmogorov-Smirnov irá ser aplicado para verificar se a distribuição que está a ser testada se ajusta à amostra, com o nível de confiança requerido.

O p-valor também permite avaliar a qualidade do ajuste (Pandis, 2013). A sua aplicação também é referida no capítulo 5.

2.4 - Conclusões

Neste capítulo foi evidenciada a importância da aplicação de métodos fiabilísticos de modo a conhecer o comportamento do equipamento ou do componente analisado como ferramenta de apoio à gestão da manutenção.

Foi referido os principais problemas práticos que geralmente estão associados à análise de dados provenientes do registo histórico de falhas de equipamentos mecânicos, nomeadamente, o reduzido número de dados e a existência de dados incompletos.

Foi também apresentada uma metodologia que permite a estimação dos parâmetros da distribuição fiabilística que melhor se ajusta aos dados obtidos.

O conhecimento do comportamento dos componentes permite reduzir significativamente os custos associados à manutenção pela otimização do número de inspeções, da diminuição do número de componentes substituídos e de stock, melhor preparação dos trabalhos de manutenção, diminuição dos tempos de reparação, por exemplo.

Capítulo 3

Seleção da distribuição

3.1 - Introdução

No capítulo 3 descreve-se com maior detalhe o passo 3 da metodologia de estimação paramétrica apresentada no capítulo 2, sendo referidos os diversos procedimentos para a seleção da distribuição que melhor se adequa para modelar um determinado conjunto de dados.

Neste capítulo, apresentam-se as noções de sistema reparável e de sistema não reparável, dando ênfase às suas diferenças e à distinta abordagem na sua aplicação nos estudos de fiabilidade.

No final do capítulo, caracteriza-se a distribuição estatística escolhida que melhor se adequa ao histórico de falhas do caso real apresentado no capítulo 5.

3.2 - Fiabilidade

A fiabilidade é definida pela norma portuguesa NP EN 13306 (2010), como a “aptidão de um bem para cumprir uma função requerida sob determinadas condições de utilização, durante um dado intervalo de tempo”.

A definição da fiabilidade tem três aspetos importantes, a função, as condições de utilização e o tempo.

A função requerida é definida como a função ou conjunto de funções consideradas como necessárias para fornecer um dado serviço.

O tempo deve ser bem definido e poderá ser tempo de funcionamento ou tempo de calendário. O tempo pode ser substituído por outras unidades como, por exemplo, ciclos, metros, energia consumida, de acordo com o sistema em análise.

As condições de utilização entendem-se como as condições operacionais e ambientais de funcionamento.

Os sistemas são projetados para realizar uma determinada função durante um período de tempo específico, tendo em conta determinadas condições de funcionamento. A alteração destas condições, pode provocar o aparecimento de falhas prematuras e consequente inaptidão na realização da função requerida.

Os modelos matemáticos aplicados na análise de fiabilidade recorrem a ferramentas de tratamento estatístico dos dados disponíveis. As variáveis aleatórias envolvidas no processo têm como domínio valores numéricos.

Considera-se que a variável aleatória é representada por uma letra maiúscula, por exemplo, T ou X e o valor numérico que a variável aleatória pode tomar é representado por uma letra minúscula. Por exemplo, se T representar o tempo de funcionamento, então t_i é o tempo referente à falha i ($i = 1, 2, \dots, n$).

A fiabilidade representa-se matematicamente pela função fiabilidade, $R(t)$, dada pela equação 3.1 (O' Connor, 2012), (Tobias, 2011),

$$R(t) = P(T > t) \quad 3.1$$

Considera-se que a variável aleatória T , não negativa, representa o tempo t . Esta equação indica a probabilidade de um equipamento não avariar antes de t .

A função fiabilidade apresenta as seguintes características:

- $R(t) \geq 0$
- $R(0) = 1$
- $R(\infty) = \lim_{t \rightarrow \infty} R(t) = 0$
- monótona decrescente

A distribuição de probabilidade de uma variável aleatória T contínua representa-se por $f(t)$ e designa-se por função densidade de probabilidade. A função densidade de probabilidade é definida de tal modo que a probabilidade do evento $P(t_1 \leq T \leq t_2)$ é igual à área sob a curva da função $f(t)$ entre $t=t_1$ e $t=t_2$, de acordo com a equação seguinte.

$$P(t_1 \leq T \leq t_2) = \int_{t_1}^{t_2} f(t).dt \quad 3.2$$

Se esta função for integrada entre o momento de entrada em funcionamento do sistema ($t_1=0$) e um momento t ($t_2=t$), obtém-se a função de probabilidade acumulada de falha, $F(t)$, conforme a equação 3.3.

$$F(t) = \int_0^t f(t).dt \quad 3.3$$

3 - Seleção da distribuição

A função de probabilidade acumulada de falha, em geral, apresenta as seguintes características:

- $F(0) = 0$
- $F(\infty) = \lim_{t \rightarrow \infty} F(t) = 1$
- monótona crescente

A função de probabilidade acumulada de falha é complementar da função fiabilidade, assim verifica-se que,

$$F(t) = 1 - R(t) \quad 3.4$$

A função fiabilidade vem expressa por,

$$R(t) = 1 - \int_0^t f(t).dt \quad 3.5$$

Assim vem,

$$R(t) = \int_t^{\infty} f(t).dt \quad 3.6$$

As funções $F(t)$ e $R(t)$ representam as áreas limitadas pela curva definida por $f(t)$ e como a área limitada por esta função é igual a 1, verifica-se que $0 \leq R(x) \leq 1$ e $0 \leq F(x) \leq 1$.

As funções fiabilidade e probabilidade acumulada de falha estão representadas na figura 3.1.

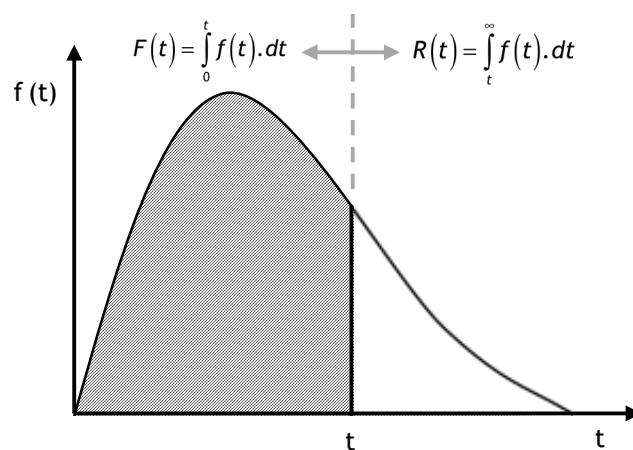


Figura 3.1 - Relação entre $f(t)$, $F(t)$ e $R(t)$.

É importante diferenciar os sistemas não reparáveis dos sistemas reparáveis. Os sistemas

não reparáveis, designados também por componentes, têm um período de vida que termina quando ocorre a primeira e única falha. Os sistemas reparáveis, também designados por equipamentos, são reparados quando ocorrem falhas e o fim da sua vida surge quando o custo de manutenção ultrapassa o custo de substituição ou quando o equipamento se torna obsoleto, sendo a alternativa de substituição mais económica.

A diferença entre componente e sistema é importante do ponto de vista da fiabilidade, na medida em que a análise da fiabilidade de cada um baseia-se em indicadores e modelos distintos que devem ser abordados de forma diferente (Pereira, 2012), (Procaccia, 2011).

3.3 - Sistemas não reparáveis

Os componentes elementares são geralmente considerados como não reparáveis. Contudo, esta avaliação depende do tipo de falha e do custo económico, ambiental e de segurança que estão associados. Um sistema não reparável ou componente após a sua primeira e única falha é substituído por um componente considerado como novo. Nesta condição, com o objetivo de realizar uma análise estatística, os novos componentes são considerados independentes e identicamente distribuídos (iid - *independent and identically distributed*). Sendo assim, os sistemas não reparáveis podem ser modelados por uma distribuição estatística, como por exemplo, a distribuição de Weibull (O' Connor, 2012), (Procaccia, 2011), (Tobias, 2011).

Consideram-se duas variáveis X e Y independentes, se e só se, para quaisquer valores x e y , verificar-se a seguinte condição,

$$P(Y = y | X = x) = P(Y = y)$$

ou equivalente,

$$P(X = x | Y = y) = P(X = x)$$

Atendendo a esta definição, a independência entre duas variáveis implica que entre elas não exista qualquer relação. Qualquer que seja o valor particular que uma das variáveis tome, não se altera a distribuição de probabilidade da outra variável.

Consideram-se dois eventos X e Y identicamente distribuídos se tiverem a mesma função de distribuição, ou seja,

$$f_x(x) = f_y(y)$$

O tempo entre o início do funcionamento de um componente novo e a sua falha é designado por tempo até à falha ou TTF (*time to failure*). Através da informação obtida em componentes iguais, submetidos a condições semelhantes, em um ou mais equipamentos, é

possível obter o tempo médio até à falha ou MTTF (*mean time to failure*). O MTTF pode ser obtido pela expressão matemática definida pelo valor esperado de uma variável,

$$MTTF = \int_0^{\infty} t \cdot f(t) dt \quad 3.7$$

A probabilidade instantânea da primeira e única falha é designada por função de risco, $h(t)$. Geralmente a função de risco é denominada por taxa instantânea de falha ou por taxa de falha da distribuição e representa a probabilidade por unidade de tempo de um componente não reparável avariar no intervalo $(t; t + \delta t)$, dado que sobreviveu até t .

Considera-se a probabilidade condicional,

$$P(t < T \leq t + \delta t | T > t) = \frac{P(t < T \leq t + \delta t)}{P(T > t)} = \frac{F(t + \delta t) - F(t)}{R(t)} = \frac{R(t) - R(t + \delta t)}{R(t)} \quad 3.8$$

Obtém-se a função de risco quando $\delta t \rightarrow 0$ na expressão 3.8.

$$\begin{aligned} h(t) &= \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t | T \geq t)}{\delta t} \\ \Leftrightarrow h(t) &= \frac{1}{R(t)} \lim_{\delta t \rightarrow 0} \frac{F(t + \delta t) - F(t)}{\delta t} = \frac{f(t)}{R(t)} \end{aligned} \quad 3.9$$

A função de risco $h(t)$ relaciona-se com a função fiabilidade, de acordo com os seguintes passos,

$$\begin{aligned} h(t) &= -\frac{dR(t)}{dt} \times \frac{1}{R(t)} \\ \Leftrightarrow \int_0^t h(t) dt &= \int_0^t -\frac{dR(t)}{R(t)} = \ln R(t) \\ \Leftrightarrow R(t) &= \exp \left[-\int_0^t h(t) dt \right] \end{aligned} \quad 3.10$$

A equação 3.10 é designada por equação fundamental da fiabilidade e pode ser utilizada em qualquer distribuição contínua. A função de risco acumulado, $H(t)$ é dada por,

$$H(t) = \int_0^t h(t) dt = \int_0^t \frac{f(t)}{R(t)} dt \quad 3.11$$

Pode-se comparar a função de risco com a função densidade de falhas, dado que a função densidade de probabilidade é a derivada em ordem a t de $F(t)$, assim,

$$f(t) \equiv \frac{\Delta F_n(t)}{\Delta t} \equiv \frac{1}{n} \frac{n(t) - n(t + \Delta t)}{\Delta t} \quad 3.12$$

em que $n(t)$ representa o número de componentes que sobrevivem até t e $n(0) = n$. Um estimador natural para a função de risco é,

$$\hat{h}(t) \equiv \frac{\hat{f}(t)}{\hat{R}(t)} \equiv \frac{\hat{f}(t)}{n(t)/n} \equiv \frac{1}{n(t)} \frac{n(t) - n(t + \Delta t)}{\Delta t} \quad 3.13$$

Assim a função de risco é dada como o número de falhas por unidade de tempo em t a dividir pelo número de componentes em risco em t .

3.4 - Sistemas reparáveis

De acordo com a NP EN 13306 (2010), define-se um bem reparável por um bem que, depois de uma falha e sob determinadas condições, pode ser repostado num estado em que poderá desempenhar a função requerida. As condições dadas podem ser económicas, ecológicas, técnicas ou outras.

Ou seja, um sistema reparável é um sistema que após cada falha é reparado, sendo reposta a sua situação operacional até uma outra possível falha. Ao contrário, um sistema não reparável após a sua primeira e única falha é substituído por um outro componente.

Considera-se por T_i , $i \geq 1$, a variável aleatória que representa o tempo acumulado de funcionamento desde o início do teste ou do arranque do sistema até à ocorrência de cada uma das falhas sucessivas e X_i , $i \geq 1$, a variável aleatória que representa o tempo entre duas falhas consecutivas, conforme indicado na figura seguinte. T_i e X_i são então variáveis que assumem valores específicos, t_i e x_i . Considera-se $t_0 = 0$ (O' Connor, 2012), (Procaccia, 2011), (Tobias, 2011).

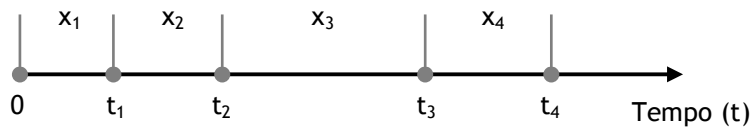


Figura 3.2 - Tempo entre falhas e tempo acumulado de funcionamento.

Assim,

$$t_i = \sum_{j=1}^i x_j \quad 3.14$$

$$x_i = t_i - t_{i-1} \quad 3.15$$

Pode-se verificar que um sistema reparável degrada-se quando o tempo entre falhas sucessivas, X_i , diminui e que, pelo contrário, se o tempo entre falhas aumenta, há uma melhoria do sistema. As atividades de manutenção podem influenciar a evolução do comportamento do sistema, por exemplo, uma operação de manutenção mal executada pode provocar a diminuição do tempo entre falhas.

Considera-se $N(t)$ uma variável aleatória que representa o número de falhas acumuladas pelo sistema ou sistemas no intervalo $(0; t)$. A média deste processo é definida por $E(N(t))$, ou seja, o valor esperado do número de falhas acumuladas pelo sistema.

Pela derivada de $E(N(t))$ em ordem ao tempo, obtém-se a taxa de ocorrência de falhas (ROCOF - Rate of occurrence of failures).

$$v(t) = \lim_{\delta t \rightarrow 0} \frac{E[N(t + \delta t) - N(t)]}{\delta t} \quad 3.16$$

O limite de equação 3.16 é o equivalente da derivada de $E(N(t))$ em relação a t .

$$v(t) = \frac{d}{dt} E(N(t)) \quad 3.17$$

O ROCOF é o limite da divisão do número médio de avarias de um sistema reparável, durante um intervalo $(t; t + \delta t)$, pela duração do intervalo de tempo δt , quando este tende para 0. Pode ser interpretada como a taxa de variação do número esperado de falhas.

O processo de sistemas reparáveis também pode ser caracterizado pela função intensidade de falha, $\lambda(t)$, definida por,

$$\lambda(t) = \lim_{\delta t \rightarrow 0} \frac{P[N(t, t + \delta t) \geq 1]}{\delta t} \quad 3.18$$

A função de intensidade de falha, $\lambda(t)$, pode ser considerada como a probabilidade da ocorrência de uma falha num pequeno intervalo $(t; t + \delta t)$, dividida pelo tamanho do intervalo δt . Assim, se $\lambda(t)$ for grande, espera-se um número maior de falhas no intervalo.

É importante diferenciar as definições da função de risco (equação 3.9) da função de intensidade de falha (equação 3.18). A função de risco $h(t)$ é a probabilidade condicionada da falha ocorrer apenas uma vez no intervalo de tempo $(t; t + \delta t)$ dividida pelo tamanho do intervalo. A função de intensidade $\lambda(t)$ é a probabilidade não condicional da ocorrência de uma falha, não necessariamente a primeira, no intervalo de tempo $(t; t + \delta t)$, dividida pelo tamanho do intervalo.

Nos sistemas reparáveis o tempo médio entre falhas, MTBF (*Mean Time Between Failures*) corresponde ao tempo médio durante o qual o equipamento permanece em

funcionamento até ocorrer uma falha, ou seja, o tempo que decorre, em média, entre duas falhas consecutivas. O valor do MTBF é determinado pelo inverso da taxa de ocorrência de falha. Se a taxa for constante o MTBF não depende do tempo.

Os sistemas ou equipamentos reparáveis têm períodos em que estão disponíveis para funcionar e períodos em que não estão disponíveis, por diferentes razões, nomeadamente por terem avariado e estarem em reparação ou por se encontrarem em manutenção. Por esse facto, a disponibilidade é uma medida que é significativamente relevante na avaliação do desempenho de um equipamento. A disponibilidade é função da maior ou menor frequência de falhas, mas também, da maior ou menor rapidez da realização das ações de manutenção, que por sua vez é dependente dos meios disponíveis e da manutibilidade do equipamento. A manutibilidade é um parâmetro de projeto que traduz a capacidade de um equipamento ser mantido em boas condições. A definição matemática mais geral da disponibilidade é a seguinte:

$$D = \frac{T_{in}}{T_{in} + T_{out}} \quad 3.19$$

T_{in} representa o período de tempo durante o qual o sistema se encontra num estado operacional (figura 3.3), podendo estar ativo ou não. T_{out} representa um período de tempo em que o sistema não está operacional e engloba o tempo de reparação (que inclui os tempos de diagnóstico, localização da avaria, de preparação da reparação, de reparação e de verificação e ensaio) e o tempo dedicado a ações de manutenção preventiva, o tempo logístico (tempo de espera por componentes e materiais para realização da ação de manutenção) e o tempo administrativo (tempo de preenchimento de impressos e de afetação do trabalho de manutenção).

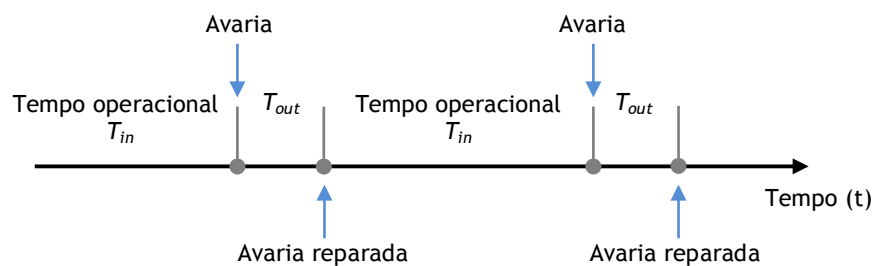


Figura 3.3 - Representação de T_{in} e T_{out} .

Quando um sistema se encontra no estado estacionário, o valor da disponibilidade pode ser obtido através dos valores do MTBF e do MTTR. O MTTR (*Mean Time To Repair*) representa o tempo médio de reparação.

$$D = \frac{MTBF}{MTBF + MTTR} \quad 3.20$$

O aumento da disponibilidade de um equipamento pode ser conseguido pela redução do número de paragens, alcançadas através de ações de manutenção.

A evolução da taxa de ocorrência de falha ao longo do tempo pode ser representada pela figura 3.4, conhecida por “curva da banheira”.

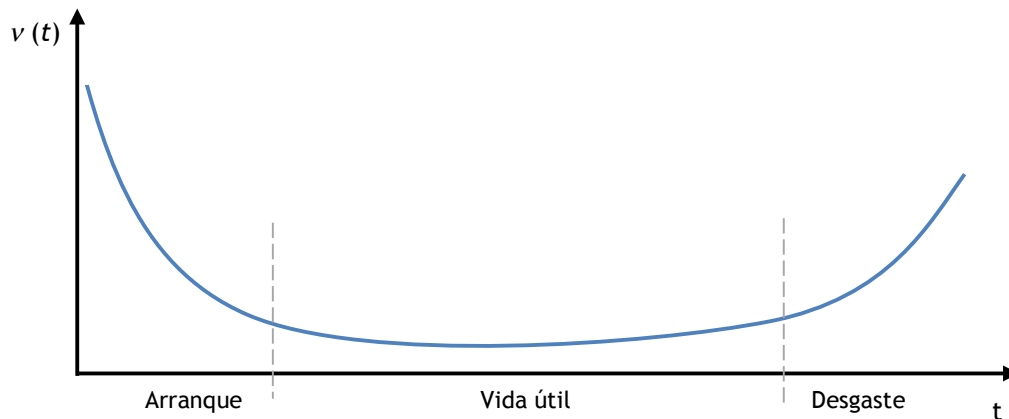


Figura 3.4 - Evolução da taxa de ocorrência com o tempo (O'Connor, 2012).

Este gráfico representa os três períodos distintos da vida do equipamento, juventude (arranque), maturidade (vida útil) e velhice (desgaste).

A fase infantil ou fase de avarias precoces é um período de tempo curto em que a taxa de ocorrência de falha é elevada, mas decrescente. Devido à execução dos mais variados testes, geralmente quando os equipamentos chegam ao cliente, esta fase foi ultrapassada.

Na fase de vida útil, a taxa de ocorrência de falha do sistema mantém-se sensivelmente constante. Se o equipamento estiver sujeito às condições para as quais foi projetado as falhas ocorrem apenas devido a causas aleatórias.

Após o período da vida útil, o equipamento entra na fase de desgaste. Nesta fase a taxa de ocorrência aumenta devido à deterioração de alguns componentes, originada por efeitos cumulativos tais como a fadiga, a corrosão ou o desgaste.

A forma da curva pode ser ligeiramente diferente em função do tipo de equipamento, por exemplo, verifica-se que os sistemas eletrónicos, em geral, têm uma longa vida útil e consequentemente é a fase mais extensa da curva enquanto que os sistemas mecânicos, geralmente, atingem mais rápido a fase de desgaste.

A curva em forma de banheira apresenta uma configuração semelhante para sistemas reparáveis e não reparáveis, contudo têm uma interpretação diferenciada. Para os sistemas reparáveis representa a taxa de ocorrência de falha em função do tempo de serviço acumulado e para os sistemas não reparáveis representa a variação da função de risco com a idade. Em termo gerais pode-se considerar uma curva da banheira para sistemas reparáveis e outra para sistemas não reparáveis.

Em sistemas reparáveis, a fiabilidade não é modelada em termos de distribuições estatísticas, mas pela utilização de um processo estocástico. O número de falhas num intervalo de tempo pode ser representado através de um processo estocástico.

Do ponto de vista matemático, um processo estocástico é definido por uma família de variáveis aleatórias, $\{X(t), t \in T\}$, definidas no conjunto T .

O conjunto T pode ser definido como um espaço de tempo. Dependendo da natureza da variável, o processo é classificado de processo com parâmetro discreto ou com parâmetro contínuo, isto é, se T é uma sequência de variáveis discretas $T=\{0, \pm 1, \pm 2, \dots\}$ ou $T=\{1, 2, \dots\}$, então o processo estocástico $\{X(t), t \in T\}$ é chamado de processo com parâmetro discreto, se T é um intervalo ou uma combinação algébrica de intervalos, por exemplo, $T=\{0 < t < +\infty\}$ ou $T=\{-\infty < t < +\infty\}$, então o processo estocástico $\{X(t), t \in T\}$ é chamado de processo estocástico com parâmetro contínuo.

Os principais modelos estocásticos aplicados a sistemas reparáveis são (Ascher, 1984), (Pereira, 2012), (Procaccia, 2011).

- processo de renovação (RP - *renewal process*)
- processo de Poisson homogéneo (HPP - *homogeneous Poisson process*)
- processo de Poisson não homogéneo (NHPP - *non-homogeneous Poisson process*).

3.4.1 - Processo de renovação

Num processo de renovação o sistema passa por uma ação de manutenção que o repõe na mesma condição inicial após a ocorrência da falha. Este processo, assume que o sistema é reposto numa condição de como novo (AGAN - *as good as new*) sempre que for reparado. A intervenção de manutenção é qualificada como uma reparação perfeita, que recoloca a intensidade de falha no valor que tinha no instante t_0 . Assim o processo de renovação não pode ser utilizado para modelar um processo de degradação. Num determinado intervalo $(0, t)$ haverá tantas renovações quantas forem as falhas. Sendo assim, o processo renovável pode ser modelado por uma distribuição estatística, como por exemplo, a distribuição de Weibull.

O processo de renovação é um processo estocástico em que $\{N(t), t \geq 0\}$.

Seja $N(t)$ o número de falhas durante o intervalo $(0, t)$. Os tempos entre falhas $x = (x_1, x_2, \dots, x_n)$ são variáveis aleatórias independentes e identicamente distribuídas (iid).

O tempo para a falha n é de

$$T_n = x_1 + x_2 + \dots + x_n \quad (n = 1, 2, \dots) \quad 3.21$$

Considerando-se que os tempos entre renovações são iid, vem que

$$P(X_n \leq t) = F(t), \quad t \geq 0, n = 0, 1, \dots \quad 3.22$$

Então, a probabilidade de o número de falhas até t ser n é dado por

$$P(N(t) = n) = P(N(t) \geq n) - P(N(t) \geq n+1) \quad 3.23$$

$$= P(T_n \leq t) - P(T_{n+1} \leq t) \quad 3.24$$

Assim a função de renovação $M(t)$ é definida como o valor esperado de número de falhas para t , isto é,

$$M(t) = E[N(t)] = \sum_{n=1}^{\infty} n \cdot P(N(t) = n) \quad 3.25$$

$$\begin{aligned} &= \sum_{n=1}^{\infty} n [P(N(t) \geq n) - P(N(t) \geq n+1)] \\ &= \sum_{n=1}^{\infty} P(N(t) \geq n) \end{aligned} \quad 3.26$$

3.4.2 - Processo de Poisson homogêneo

No processo de Poisson homogêneo considera-se que o sistema tem início de funcionamento em $t=0$ e funciona durante um período de tempo t . O número de falhas $N(t)$ a que o sistema está sujeito durante este período e os tempos entre falhas sucessivas são aleatórios. Durante o funcionamento do sistema, os tempos entre falhas sucessivas, X_i ($i = 1, 2, \dots, n$) são variáveis independentes e são distribuídas exponencialmente com intensidade de falhas λ constante. Este processo é considerado como independente e identicamente distribuído (iid).

Pode-se caracterizar o processo de Poisson por:

- $N(0) = 0$;
- $N(t)$ tem incrementos independentes, ou seja, se (a, b) e (c, d) forem intervalos disjuntos, então $N(a, b)$ e $N(c, d)$ são independentes;
- A função intensidade de falha do processo é dada por:

$$\lambda(t) = \lim_{\delta t \rightarrow 0} \frac{P(N(t, t + \delta t) = 1)}{\delta t} \quad 3.27$$

Esta característica implica a ocorrência máxima de uma falha em δt .

- O processo é regular, ou seja:

$$\lim_{\delta t \rightarrow 0} \frac{P(N(t, t + \delta t) \geq 2)}{\delta t} = 0 \quad 3.28$$

Esta característica elimina a probabilidade de ocorrência de mais do que uma falha em δt .

O processo de Poisson homogêneo pode ser visto como um caso particular do processo de renovação.

Seja um processo de Poisson homogêneo, com tempos de falhas $0 < t_1 < t_2 < \dots < t_n$ e sejam $x_i = t_i - t_{i-1}$ ($i = 1, 2, \dots, n$), com $t_0 = 0$, os tempos entre falhas. Considera-se que os tempos entre falhas são independentes, com distribuição exponencial, com média $1/\lambda$, tal que,

$$F(t) = 1 - \exp(-\lambda t) \quad 3.29$$

A função de probabilidade conjunta de T_1, T_2, \dots, T_n é da forma (Pereira, 2012),

$$f(t_1, t_2, \dots, t_n) = f_1(t_1) f_2(t_2 | t_1) \dots f_n(t_n | t_1, t_2, \dots, t_{n-1}) \quad 3.30$$

$$f(t_1, t_2, \dots, t_n) = \lambda^n \exp(-\lambda t_n) \text{ com } t_n > t_{n-1} > \dots > t_2 > t_1 > 0 \quad 3.31$$

Assim em geral, a função fiabilidade condicional de T_n é dada por,

$$R(t_n | t_1, t_2, \dots, t_{n-1}) = \exp\left(-\int_{t_{n-1}}^{t_n} \lambda dt\right) = \exp[-\lambda(t_n - t_{n-1})] \text{ com } t_n > t_{n-1} > 0 \quad 3.32$$

3.4.3 - Processo de Poisson não homogêneo

Num sistema complexo composto por vários módulos reparáveis e não reparáveis, a maior parte das intervenções de manutenção implicam a substituição de apenas um número reduzido de elementos do sistema. Esses elementos são considerados como novos mas a maioria dos outros elementos do sistema continuam no mesmo estado depois da reparação.

Em consequência deste procedimento, a fiabilidade do sistema não é efetivamente alterada pela substituição de um número reduzido de elementos e praticamente mantêm o mesmo nível de fiabilidade que tinha antes da falha.

Assim o sistema é considerado numa condição de tão mal quanto antes (ABAO - *as bad as old*) em que cada reparação repõe o estado do equipamento para quase o estado que tinha antes da falha. A intervenção de manutenção pode ser qualificada como reparação mínima.

O processo de Poisson não homogêneo é um processo de Poisson cuja função intensidade de falha não é constante. Neste processo as variáveis aleatórias entre falhas sucessivas não têm a mesma distribuição, assim, o processo não é nem independente nem identicamente distribuído.

No processo de Poisson não homogêneo a taxa acumulada num intervalo $(0, t)$ é dada por,

$$\Lambda(t) = \int_0^t \lambda(t) dt \quad 3.33$$

Assim o número de falhas no intervalo de tempo $(0, t)$ é dado pela seguinte equação,

$$P(N(t) = n) = \frac{(\Lambda(t))^n}{n!} \exp(-\Lambda(t)) \quad 3.34$$

O número de falhas num intervalo (t_1, t_2) , vem,

$$P(N(t_2) - N(t_1) = n) = \frac{\left(\int_{t_1}^{t_2} \lambda(t) dt \right)^n}{n!} \exp\left(- \int_{t_1}^{t_2} \lambda(t) dt \right) \quad 3.35$$

A função de intensidade para o processo de Poisson não homogêneo, pode ser representada por vários modelos, dos quais se podem destacar os seguintes,

- Função Power Law: $\lambda(t) = \eta \cdot \beta \cdot t^{\beta-1}$ 3.36

em que η e β são os parâmetros do modelo.

- Função log-linear: $\lambda(t) = \exp(\alpha + \beta t)$ 3.37

em que η e β são os parâmetros do modelo, $\eta, \beta > 0$.

A figura seguinte representa três situações que podem ocorrer em sistemas em relação ao tempo entre falhas.

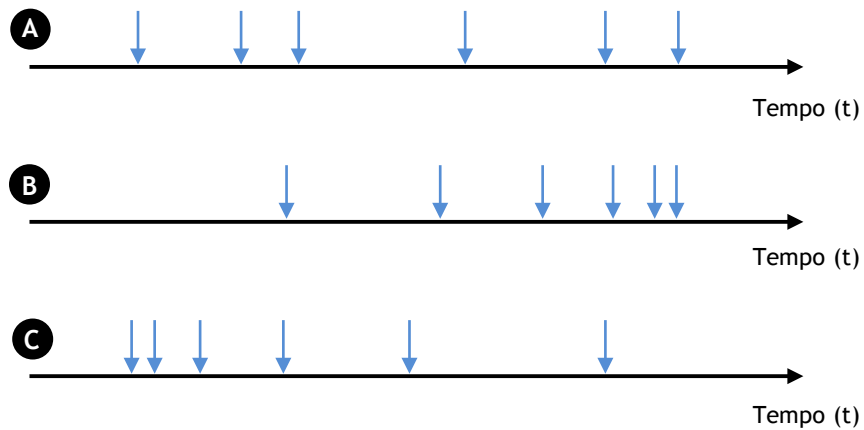


Figura 3.5 - Possível tendência do tempo entre falhas.

Pode-se verificar que as três situações, A, B e C, representadas na figura, são bastante diferentes entre si. No caso A, não se consegue verificar uma tendência notória a partir dos tempos de falha representados, assim pode ser feita uma suposição de condição iid, dado que o tempo entre as falhas é aparentemente independente da idade do equipamento. Contudo, para B e C está claramente presente uma tendência. Em B é evidente uma diminuição do

tempo entre falhas enquanto em C ocorre o seu aumento. Sempre que ocorrem estas duas últimas situações e há uma significativa evidência que esteja a ocorrer um processo de envelhecimento, a usual abordagem do processo de renovação não pode ser utilizada e uma alternativa não estacionária deve ser utilizada para modelar o tempo entre falhas do sistema. Note-se que a idade do equipamento, refere-se à idade do sistema em análise, medido a partir do momento em que foi colocado em funcionamento, em oposição ao tempo decorrido desde a última reparação que é utilizado no processo de renovação.

3.5 - Análise de tendência

Do subcapítulo anterior, verifica-se uma clara necessidade na análise da tendência do tempo entre falhas. Assim, um primeiro passo na seleção do modelo que mais se adequa ao sistema deve ser a avaliação da existência de tendência ou dependência do tempo. Existem várias técnicas para realizar essa tarefa, contudo apresentam-se as mais importantes (Ascher, 1984), (O' Connor, 2012), (Procaccia, 2011), (Vaurio, 1998).

Antes de apresentar as técnicas de teste de tendência é importante identificar os possíveis padrões que se podem encontrar quando se analisa os dados. O padrão da tendência pode ser monótona ou não monótona.

No caso de ser uma tendência monótona, tal como representado na figura 3.5, considera-se que o sistema está a melhorar o seu desempenho se o tempo entre as falhas tende a ficar maior (tendência decrescente no número de falhas) e considera-se que o sistema está a deteriorar-se se o tempo entre as falhas tende a ficar mais reduzido (tendência crescente no número de falhas).

A tendência é não monótona quando a tendência muda com o tempo ou quando se repete em ciclos.

3.5.1 - Métodos gráficos

3.5.1.1 - Gráfico do número acumulado de falhas

O método mais simples para o teste de tendência é a representação gráfica do número acumulado de falhas em função do tempo. Quando o gráfico resulta numa linha reta, pode ser assumido que não existe tendência e é aceite a mesma distribuição do tempo entre falhas.

A figura 3.6 ilustra possíveis gráficos que podem ser esperados.

O gráfico A ilustra, claramente, a existência de uma tendência nos dados, enquanto que no gráfico B não é evidente uma tendência. Pode surgir um gráfico semelhante ao gráfico C, onde, em vez de uma tendência notória, duas ou mais linhas retas podem ser representadas. Este comportamento pode ser devido a uma mudança na política de manutenção ou devido à alteração das condições operacionais do equipamento. Quando surge esta situação, uma

alternativa é não considerar os dados não representativos da situação atual e assim pode resultar num gráfico sem tendência e consequente suposição de processo de renovação. Quando ocorre um gráfico semelhante ao gráfico D, pode estar presente uma tendência não monótona.

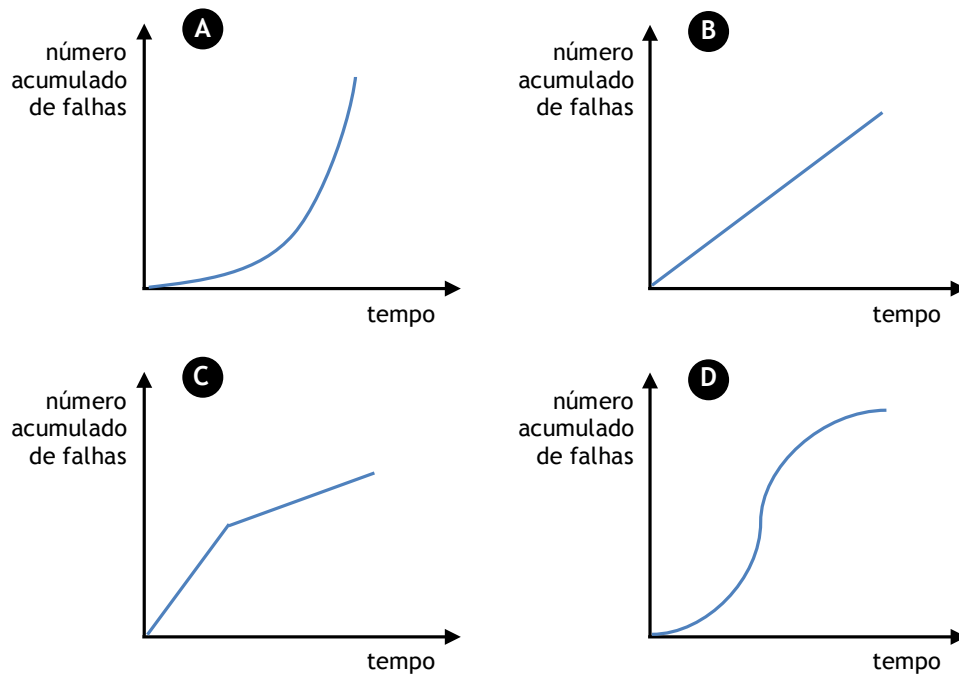


Figura 3.6 - Número acumulado de falhas em função do tempo t (A- tendência crescente; B- sem tendência; C- dois períodos diferentes; D- tendência não monótona).

Este tipo de teste é muito simples de executar, não necessita de quaisquer cálculos e é muito poderoso quando há evidente tendência nos dados. Quando está presente uma ligeira tendência, esta solução pode não ser suficiente e deve ser realizada uma análise analítica. A fraqueza deste teste é que a avaliação da tendência baseia-se na interpretação do gráfico obtido a partir dos dados.

3.5.1.2 - Gráfico de Nelson-Aalen

Outro teste gráfico de tendência útil é o gráfico de Nelson-Aalen. Este teste utiliza um estimador não paramétrico da função intensidade acumulada de falha. O estimador é dado por,

$$\hat{\Lambda}(t) = \sum_{T_{ij} \leq t} \frac{1}{Y(T_{ij})} \quad 3.38$$

Onde T_{ij} é o tempo da i -ésima falha do j -ésimo sistema em análise, $Y(T_{ij})$ é o número de sistemas em funcionamento imediatamente antes do tempo T_{ij} . Considera-se $\Lambda(t)=0$ para $t < \min(T_{ij})$. A equação 3.38 é válida para múltiplos sistemas em análise, $j=1, 2, \dots, m$.

Se não há uma tendência, o gráfico apresenta-se linear e qualquer desvio de uma reta indica algum tipo de tendência. Se o gráfico apresentar uma concavidade para baixo indica que o tempo entre falhas aumenta com o tempo. Se o gráfico apresentar uma concavidade para cima indica que o tempo entre falhas diminui com o tempo.

Quando apenas um sistema é observado então o gráfico de Nelson-Aalen é equivalente ao gráfico do número acumulado de falhas.

3.5.1.3 - Gráfico TTT (tempo total em teste)

Consideram-se m sistemas independentes com a mesma função intensidade de falha com o intervalo de tempo de observação $(0, s)$. O número total de falhas é

$$N = \sum_{i=1}^m n_i \quad 3.39$$

n_i é o número de falhas observadas para cada sistema.

Seja $p(u)$ o número de sistemas sobreviventes até ao instante u . Se todos os sistemas sobreviverem até s , $p(u)$ é igual a m . Então,

$$T(t) = \int_0^t p(u) du \quad 3.40$$

representa o tempo total em teste (*total time on test*) até ao instante s .

O quociente

$$\frac{T(S_k)}{T(S)} = \frac{\int_0^{s_k} p(u) du}{\int_0^s p(u) du} \quad 3.41$$

é designado de tempo total em teste escalado (*scaled total time on test*). S_k é o tempo do k -ésimo tempo de falha.

O gráfico com abcissa k/N e ordenada $T(S_k)/T(S)$, $k = 1, \dots, N$ é designado de gráfico TTT.

3.5.2 - Métodos analíticos

3.5.2.1 - Teste de Mann

Neste teste a hipótese nula é o processo de renovação e a hipótese alternativa é não ser um processo de renovação (Vaurio, 1998).

O teste de Mann compara cada tempo entre falhas com a anterior e conta o número de ordens inversas, M . A ordem é considerada inversa sempre que ocorre $X_i < X_j$ para $i < j$. Ou seja,

$$M = \sum_{i=1}^{n-1} \sum_{j=i+1}^n I(t_i < t_j) \quad 3.42$$

em que $I(.)$ é um indicador variável usado para contar o número de ordens inversas presente no conjunto de dados, que assume o valor de 1 sempre que $t_i < t_j$. A hipótese nula é aceite se M for igual a $n(n-1)/4$, em que n é o número de tempos entre falhas. Se M estiver afastado desse valor indica a presença de tendência nos dados observados.

3.5.2.2 - Teste de Laplace

O teste de Laplace pretende verificar se um conjunto de dados que fazem parte de um processo de ocorrências aleatórias são independentes e identicamente distribuídas (iid) e assim constituem um processo de Poisson homogêneo com uma intensidade de falhas λ constante.

Especifica-se a hipótese nula, H_0 , em confronto com a hipótese alternativa, H_1 , do seguinte modo,

H_0 : HPP;

H_1 : NHPP.

Se a hipótese nula não é rejeitada, então pode-se assumir que o tempo entre falhas é independente e identicamente distribuído. Se for rejeitada é considerada a condição NHPP.

Para verificar estatisticamente a veracidade da hipótese nula é necessário calcular o valor da estatística de teste, L , pela equação seguinte,

$$L = \frac{\sum_{j=1}^{\hat{n}} T_j - \frac{1}{2} \hat{n}(b+a)}{\sqrt{\frac{1}{12} \hat{n}(b-a)^2}} \quad 3.43$$

onde T_j é o tempo decorrido até à j -ésima falha, (a, b) é o intervalo de tempo da observação e \hat{n} é o número de falhas e é dado por,

$$\hat{n} = \begin{cases} n & \text{(se o teste é limitado pelo tempo)} \\ (n-1) & \text{(se o teste é limitado pelo número de falhas)} \end{cases}$$

A equação 3.43 é aplicada para situações apenas quando um sistema é analisado. A generalização do teste de Laplace para m sistemas é dado pela seguinte equação,

$$L = \frac{\sum_{i=1}^m \sum_{j=1}^{\hat{n}_i} T_{ij} - \sum_{i=1}^m \frac{1}{12} \hat{n}_i (b_i - a_i)}{\sqrt{\frac{1}{12} \sum_{i=1}^m \hat{n}_i (b_i - a_i)^2}} \quad 3.44$$

O critério de rejeição é dado pela seguinte condição,

$$L < -z_{\alpha/2} \text{ ou } L > z_{\alpha/2}$$

em que α é o nível de significância e $z_{\alpha/2}$ é o valor da normal padronizado referente ao nível de significância.

Sempre que se verificar esta condição rejeita-se a hipótese nula e o teste diz-se conclusivo.

Se L estiver dentro do seguinte intervalo,

$$-z_{\alpha/2} \leq L \leq z_{\alpha/2}$$

aceita-se a hipótese nula e o teste diz-se inconclusivo.

3.5.2.3 - Teste de Lewis-Robinson

Este teste é utilizado para testar a condição de processo de renovação. Neste teste a hipótese nula é o processo de renovação e a hipótese alternativa é não ser um processo de renovação (Vaurio, 1998).

O valor da estatística de teste, LR , é obtido pela divisão do valor da estatística de teste de Laplace, L , pelo valor estimado do coeficiente da variação do tempo entre falhas, \hat{C} , que é calculado pela seguinte equação,

$$\hat{C} = \frac{\sigma(x)}{\bar{x}} \quad 3.45$$

onde x é uma variável aleatória que representa o tempo entre falhas do sistema. Assim o valor da estatística de teste, LR , é dado por,

$$LR = \frac{L}{\hat{C}} \quad 3.46$$

o valor de L é determinado pela equação 3.43.

Do mesmo modo do que o teste de Laplace, o critério de rejeição é dado pela seguinte condição,

$$LR < -z_{\alpha/2} \text{ ou } LR > z_{\alpha/2}$$

Sempre que se verificar esta condição rejeita-se a hipótese nula e o teste diz-se conclusivo.

Se LR estiver dentro do seguinte intervalo,

$$-z_{\alpha/2} \leq LR \leq z_{\alpha/2}$$

aceita-se a hipótese nula e o teste diz-se inconclusivo.

3.5.2.4 - Teste MIL-HDBK

Do mesmo modo que no teste de Laplace, a hipótese nula é referente ao processo de Poisson homogéneo e a hipótese alternativa é referente ao processo de Poisson não homogéneo. O valor da estatística de teste, MH , é distribuído segundo a distribuição do Qui-quadrado com $2\hat{n}$ graus de liberdade. Para um sistema é definido por,

$$MH = 2 \sum_{j=1}^{\hat{n}} \ln \left(\frac{b-a}{T_j-a} \right) \quad 3.47$$

onde a , b , T_j e \hat{n} têm o mesmo significado do que no teste de Laplace. A generalização do teste para m sistemas é dado pela seguinte equação,

$$MH = 2 \sum_{i=1}^m \sum_{j=1}^{\hat{n}_i} \ln \left(\frac{b_i - a_i}{T_{ij} - a_i} \right) \quad 3.48$$

3.6 - Modelo de processo de seleção

A figura 3.7 apresenta um processo com base nas características dos sistemas reparáveis e nos métodos referidos anteriormente que permite determinar um modelo que se ajusta aos dados do sistema (Vaurio, 1998).

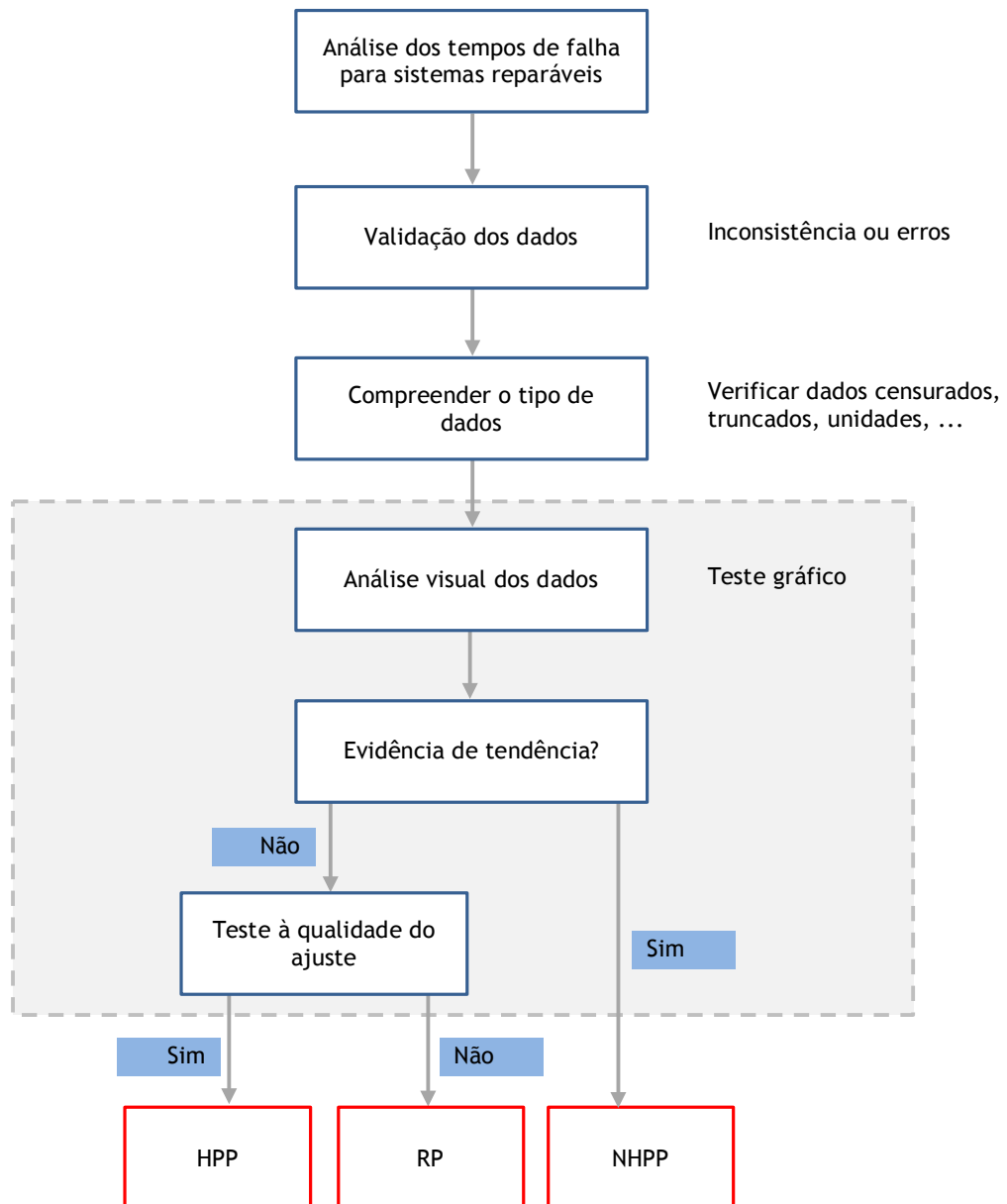


Figura 3.7 - Modelo de processo de seleção para sistemas reparáveis (adaptado de Vaurio, 1998).

Os testes de tendência utilizados, conforme referido anteriormente, podem ser testes gráficos ou analíticos, que são aplicados conforme os dados do sistema. Os testes gráficos são simples de utilizar e permitem uma visualização gráfica do resultado. Quando o resultado é inconclusivo utilizam-se os testes analíticos.

3.7 - Caracterização da distribuição selecionada

No capítulo 5 são analisados componentes mecânicos não reparáveis, na presença de dados completos e censurados à direita. Perante este contexto o procedimento utilizado para determinar a distribuição que melhor se ajusta aos dados foi pelo conhecimento do comportamento do componente em análise e de outros componentes mecânicos com comportamento semelhante e que foram alvo de estudo em outros trabalhos (Abernethy, 2006), (Ojile, 2010), (Rinne, 2009), (Wang, 2004).

Assim, a distribuição escolhida que, presumivelmente, melhor se ajusta aos dados é a distribuição de Weibull devido à sua grande flexibilidade, isto é, engloba funções com intensidade de falha constantes, crescentes e decrescentes, dependendo do valor do parâmetro de forma (Rinne, 2009), (Tobias, 2011).

3.7.1 - Distribuição de Weibull

A distribuição de Weibull tem início em 1928, quando os investigadores, Fisher e Tippet, deduziam esta distribuição de extremo valor. No final de 1930, um professor Sueco, Waloddi Weibull, derivou a mesma distribuição e a publicação do seu trabalho em 1951 tornou esta distribuição conhecida (Abernethy, 2006).

Hoje, a distribuição de Weibull tem uma larga aplicação em diversas áreas. Estas aplicações incluem o uso da distribuição para modelar a velocidade do vento, da chuva, das enchentes ou terremotos, a idade de início da doença, resistência dos materiais, etc. No entanto, o uso mais amplo da distribuição é em estudos de fiabilidade, onde a distribuição Weibull tem provado ser satisfatória na modelação dos fenómenos de fadiga e de vida de muitos dispositivos, como por exemplo, rolamentos, veios e motores (Abernethy, 2006), (McCool, 2012), (Murthy, 2004), (Rinne, 2009). Devido à sua ampla aplicação em estudos de fiabilidade, a análise de dados, é frequentemente chamada de análise de Weibull (Wang, 2004).

A popularidade da distribuição de Weibull deve-se à sua grande flexibilidade, isto é, engloba funções com intensidade de falha constantes, crescentes e decrescentes, dependendo do valor do parâmetro de forma (Rinne, 2009), (Tobias, 2011), como se verifica nos pontos seguintes onde se descreve e caracteriza a distribuição.

A forma geral da distribuição de Weibull tem três parâmetros, o parâmetro de escala, η , o parâmetro de forma, β e o parâmetro de localização, γ . A função densidade de probabilidade é dada por,

$$f(t) = \frac{\beta}{\eta} \left(\frac{t-\gamma}{\eta} \right)^{\beta-1} \exp \left[- \left(\frac{t-\gamma}{\eta} \right)^{\beta} \right], \quad t \geq \gamma \quad 3.49$$

Na análise dos dados de fiabilidade, o parâmetro de localização é frequentemente omitido. O parâmetro de localização diferente de zero não deve ser usado a menos que haja

uma justificação física para um período de tempo com uma probabilidade de falha igual a zero (Dodson, 2006).

Constata-se não haver uma convenção consistente na literatura para a representação dos parâmetros da distribuição de Weibull. Inicialmente o parâmetro de escala e o parâmetro de forma, eram representados pela letra c e m , respetivamente. Esta tese representa o parâmetro de escala por η e o parâmetro de forma por β conforme utilizado por alguns autores de referência (Abernethy, 2006), (Meeker, 1998), (O'Connor, 2012). Alguns autores também representam o parâmetro de forma por β mas o parâmetro de escala é representado por α (Nelson, 1982).

Esta tese centra-se nos métodos de estimação de parâmetros para a distribuição de Weibull de dois parâmetros. Salvo indicação em contrário, a distribuição de Weibull nesta tese refere-se à distribuição de Weibull de dois parâmetros.

Os dados necessários para a determinação da fiabilidade do sistema podem ser obtidos a partir de testes experimentais ou no decurso da sua utilização conforme referido no capítulo anterior.

A função densidade de probabilidade, para a distribuição de Weibull de dois parâmetros é dada por,

$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} \exp \left[- \left(\frac{t}{\eta} \right)^{\beta} \right] \quad 3.50$$

onde o parâmetro de escala, η e o parâmetro de forma, β , assumem valores positivos.

A partir das equações 3.3 e 3.50 é possível obter a função de probabilidade acumulada que é dada por,

$$F(t) = 1 - \exp \left[- \left(\frac{t}{\eta} \right)^{\beta} \right] \quad 3.51$$

No contexto da fiabilidade, $F(t)$ é a probabilidade que uma unidade aleatória da população em análise tem em falhar no tempo t ($t > 0$), ou a fração das unidades da população que falham no tempo t (Tobias, 2011).

Conforme indicado na equação 3.4, o complemento de $F(t)$ é a função de fiabilidade $R(t)$, ou seja,

$$R(t) = \exp \left[- \left(\frac{t}{\eta} \right)^{\beta} \right] \quad 3.52$$

Outra expressão habitualmente associada à fiabilidade é o tempo médio até à falha (*mean time to failure* - MTTF),

$$MTTF = \eta \cdot \Gamma\left(1 + \frac{1}{\beta}\right) \quad 3.53$$

onde Γ representa a função Gama.

A função intensidade de falha, $\lambda(t)$, é dada por,

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \quad 3.54$$

O percentil do tempo de vida vem,

$$t_p = \eta \left[-\ln(1-p) \right]^{\frac{1}{\beta}} \quad 3.55$$

Todas as funções anteriormente referidas são em função dos dois parâmetros da distribuição de Weibull.

O efeito do parâmetro de escala e de forma sobre a distribuição de Weibull é descrito separadamente.

a) Parâmetro de escala, η

A figura 3.8 apresenta a função densidade de probabilidade da distribuição de Weibull com diferentes valores de η e com um valor comum de β ($\beta = 3$).

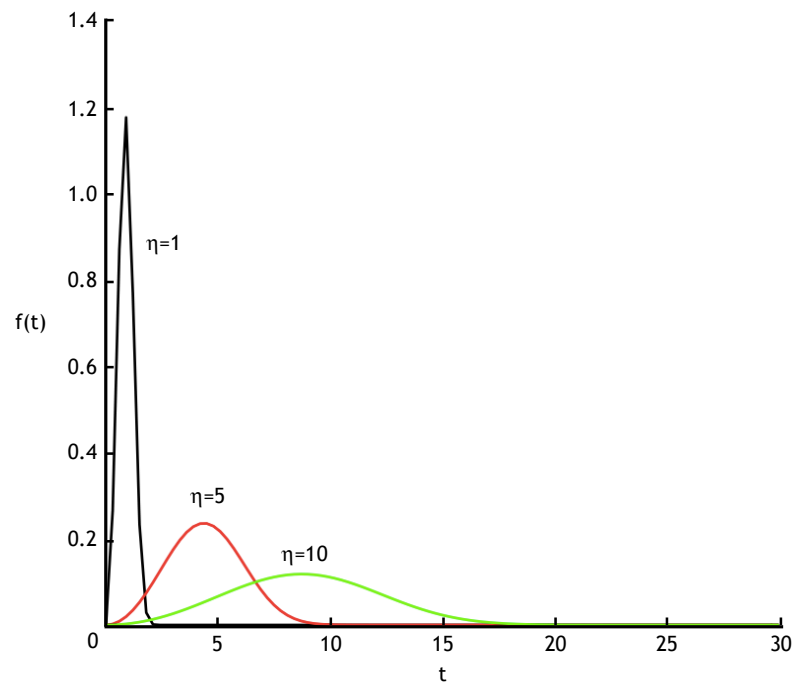


Figura 3.8 - O efeito de η na função densidade de probabilidade da distribuição de Weibull ($\beta = 3$).

Como se pode observar o aumento ou diminuição de η , enquanto β é mantido inalterado, tem um efeito de esticar a distribuição para a direita ou pressionar a distribuição para a esquerda e não tem qualquer efeito sobre a forma da distribuição. De facto, a alteração do parâmetro de escala é o mesmo que uma mudança de escala das abcissas.

O parâmetro η tem a mesma unidade que t , ou seja, horas, ciclos, etc.

Da equação 3.55, quando $p = 0,632$, obtém-se

$$t_{0,632} = \eta \quad 3.56$$

Assim η é o momento em que 63,2% da população falha. O parâmetro η é frequentemente chamado de vida característica.

b) Parâmetro de forma, β

O parâmetro de forma β é de grande importância para a distribuição de Weibull, pois determina a forma da função densidade de probabilidade e caracteriza a tendência da função intensidade de falha.

A figura 3.9 ilustra vários exemplos típicos da função densidade de probabilidade com diferentes valores de β e η fixo.

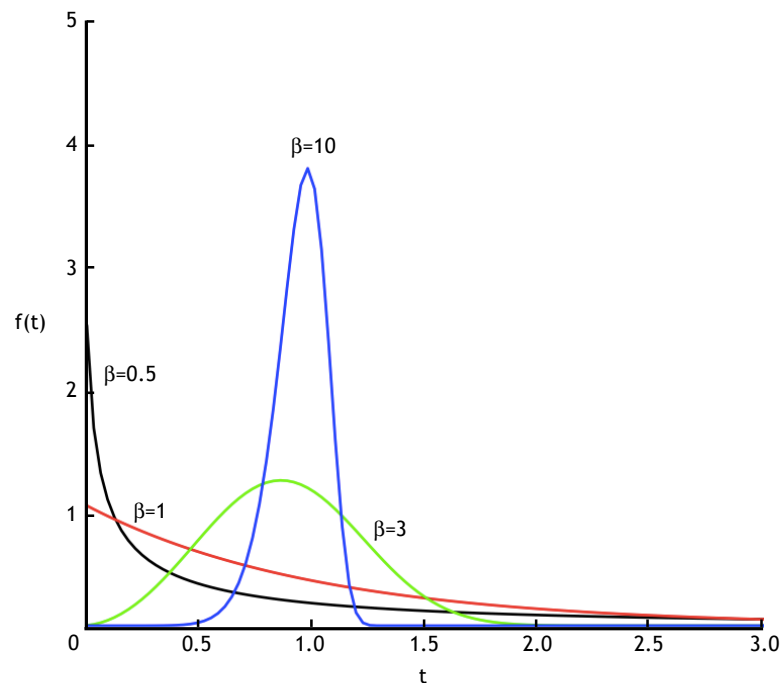


Figura 3.9 - O efeito de β na função densidade de probabilidade da distribuição de Weibull ($\eta = 1$).

Pode-se observar na figura 3.9 que, quando $0 < \beta < 1$, a função densidade de probabilidade é exponencialmente decrescente. Para $\beta = 1$, a distribuição de Weibull reduz-se à distribuição exponencial. Quando $\beta > 1$, a função densidade de probabilidade está inclinada para a direita. Quando $3 \leq \beta \leq 4$, a função densidade de probabilidade tem aproximadamente a forma de um sino, que é semelhante à distribuição normal.

A figura 3.10 ilustra o efeito de β na função intensidade de falha com η fixo.

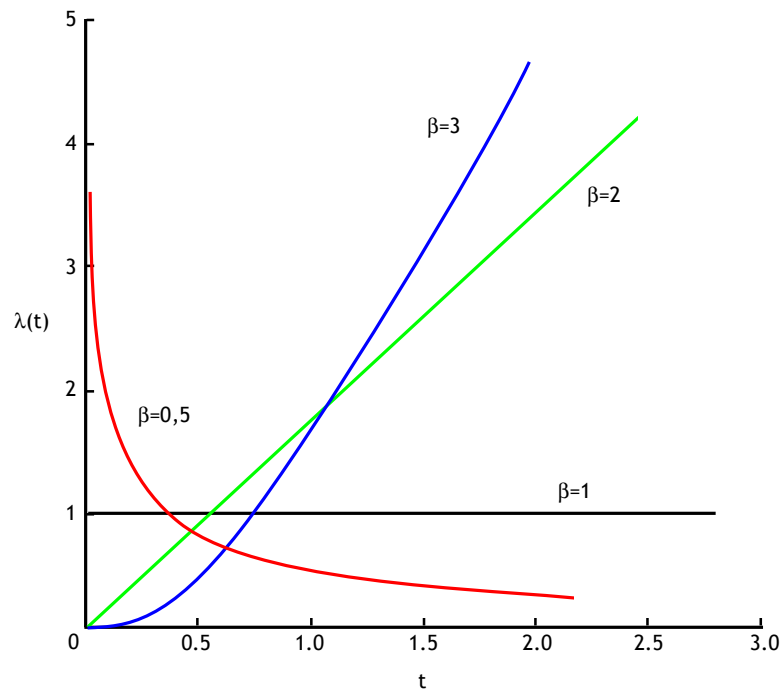


Figura 3.10 - O efeito de β na função intensidade de falha ($\eta = 1$).

A figura 3.10 ilustra a relação entre β e a função intensidade de falha. Como se pode observar, quando $0 < \beta < 1$, a função intensidade de falha é exponencialmente decrescente (idêntico à função densidade de probabilidade). Para $\beta = 1$, a função intensidade de falha é constante e igual a $\lambda(t) = 1/\eta$. Quando $\beta > 1$, a função intensidade de falha é monótona crescente. Um caso especial é quando $\beta = 2$, onde a função intensidade de falha é linearmente crescente. A distribuição neste caso é chamada distribuição de Rayleigh.

Nos outros casos a função intensidade de falha aumenta com taxas diferentes. A tabela 3.1 resume as características típicas da função densidade de probabilidade e da função intensidade de falha em função da variação de β .

Tabela 3.1 - Características típicas da função densidade de probabilidade e da função intensidade de falha em função da variação de β .

| Parâmetro de forma | Função densidade de probabilidade | Função intensidade de falha |
|-----------------------|--|------------------------------|
| $0 < \beta < 1$ | Exponencialmente decrescente do infinito | Exponencialmente decrescente |
| $\beta = 1$ | Exponencialmente decrescente de $1/\alpha$ | Constante |
| $\beta > 1$ | Sobe para o máximo e depois diminui | Crescente |
| $\beta = 2$ | Caso especial - Distribuição de Rayleigh | Linearmente crescente |
| $3 \leq \beta \leq 4$ | Forma de um sino, semelhante à distribuição normal | Rápido crescimento |

3.8 - Conclusões

A descrição apresentada dos sistemas reparáveis e não reparáveis permite determinar qual se enquadra melhor no sistema em estudo e assim especificar o processo de análise mais adequado.

Neste capítulo também é abordada a caracterização da distribuição de Weibull que serve de base para a aplicação dos métodos de estimação referidos no capítulo 4.

Capítulo 4

Estimação dos parâmetros da distribuição de Weibull

4.1 - Introdução

Neste capítulo apresenta-se os conceitos e fundamentos teóricos de alguns dos métodos de estimação dos parâmetros da distribuição de Weibull, divididos em dois grupos, métodos gráficos e métodos analíticos. Os métodos analíticos utilizados para estimar os parâmetros de Weibull são, o método dos mínimos quadrados e o método de máxima verosimilhança. É aplicado a simulação de Monte Carlo para comparar o desempenho dos vários métodos.

Como as equações de máxima verosimilhança em muitas situações não apresentam solução analítica, para determinar as suas soluções, apresentam-se dois métodos numéricos de optimização para a sua resolução, designadamente o método de Newton-Raphson e o algoritmo Expectation-Maximization (EM).

A análise do algoritmo EM é feita para dados completos e incompletos. Contudo, é dado maior ênfase na presença de dados censurados à direita tipo 1, porque será necessário a sua aplicação para o caso em estudo, que é apresentado no capítulo 5.

4.2 - Métodos de estimação dos parâmetros da distribuição de Weibull

Desde que a distribuição de Weibull se tornou amplamente reconhecida na década de 50, vários métodos foram propostos para estimar os seus parâmetros (Ahmed, 2013), (Akram, 2014), (Birolini, 2014), (O'Connor, 2012), (Procaccia, 2011), (Teimouri, 2015), (Wang, 2014).

Os métodos de estimação dos parâmetros da distribuição de Weibull são normalmente divididos em dois grupos: métodos gráficos e métodos analíticos. Os métodos Bayesianos são

outra metodologia a ser considerada, contudo não são alvo de estudo neste trabalho (Ahmed, 2014), (Guure, 2012), Procaccia, 2011).

Este capítulo fornece uma visão geral dos métodos de estimação dos parâmetros da distribuição de Weibull. É impossível listar todos os trabalhos apresentados na literatura, assim, o foco é dado aos métodos mais relevantes.

4.2.1 - Método gráfico de estimação

O método gráfico mais utilizado para a estimação dos parâmetros da distribuição de Weibull é o método gráfico de probabilidade de Weibull (Abernethy, 2006).

O gráfico de probabilidade de Weibull (*Weibull probability plotting* - WPP) foi proposto em 1959 por John K. Kao e tem sido o tradicional método gráfico para estimar os parâmetros de Weibull. Apesar de ser menos utilizado nos dias de hoje, a sua compreensão fornece uma boa base para a utilização de ferramentas de software específico. Além disso, a maioria dos softwares comercialmente disponíveis, que permitem efetuar a análise da distribuição de Weibull, usam este formato gráfico. Para além de possibilitar estimar os parâmetros de um modo simples, permite a obtenção de uma confirmação visual rápida do ajustamento de um determinado conjunto de dados e a identificação de pressupostos errados, o que é muito importante em qualquer análise de dados.

O gráfico de probabilidade de Weibull baseia-se na linearização da função de probabilidade acumulada. A linearização da função de probabilidade acumulada de Weibull é obtida ao logaritmar duas vezes ambos os lados da equação 3.51. A equação 3.51 pode ser re-escrita da seguinte forma (Abernethy, 2006), (Meeker, 1998), (O'Connor, 2012),

$$1 - F(t) = \exp \left[- \left(\frac{t}{\eta} \right)^\beta \right] \quad 4.1$$

Ao logaritmar duas vezes ambos os lados da equação 4.1, obtém-se a seguinte expressão

$$\begin{aligned} \ln[1 - F(t)] &= - \left(\frac{t}{\eta} \right)^\beta \\ \ln[-\ln(1 - F(t))] &= \beta \ln(t) - \beta \ln(\eta) \end{aligned} \quad 4.2$$

admitindo que,

$$\begin{aligned} X &= \ln(t), \\ Y &= \ln[-\ln(1 - F(t))], \end{aligned}$$

$$B = -\beta \ln(\eta).$$

A expressão 4.2 resulta numa equação mais simples,

$$Y = \beta X + B \quad 4.3$$

O gráfico de probabilidade de Weibull é dimensionado com base na equação 4.2 e deste modo apresenta uma reta quando a distribuição de Weibull se ajusta aos dados da amostra. O eixo X representa as observações t (por exemplo, tempos de falhas) obtidas de ensaios ou no contexto operacional (em fase de exploração). O eixo Y representa a função de probabilidade acumulada de falha $F(t)$ para cada tempo de falha.

A reta que melhor se ajusta aos pontos é traçada a “olho”. Este procedimento relativamente subjetivo poderá traduzir-se no enviesamento dos resultados obtidos. Erro esse que não é quantificável.

O valor de $F(t)$ para o tempo de falha é desconhecido sem o valor de η e β , portanto, só pode ser estimado.

4.2.1.1 - Estimação de $F(t)$

A estimação de $F(t)$ é frequentemente referida pela determinação do valor da posição no eixo Y no gráfico de probabilidade de Weibull.

A escolha do método para obter o valor de $F(t)$, depende se os dados são completos ou censurados. Os fundamentos teóricos usualmente mencionados nas referências bibliográficas para a determinação do valor de $F(t)$ para dados completos e censurados são resumidamente apresentados (Fothergill, 1990), (Nelson, 1982), (Olteanu, 2010), (Rinne, 2009).

a) Estimação de $F(t)$ para dados completos

A forma geral dos estimadores para dados completos pode ser expressa pela seguinte expressão:

$$\hat{F}(t) = \frac{i - c_1}{n + c_2} \quad 4.4$$

Sendo c_1 e c_2 números reais.

A média e a mediana são habitualmente utilizadas para a determinação do valor de $F(t)$. A média apresenta uma forma simples,

$$\hat{F}_{(i)} = \frac{i}{n + 1} \quad 4.5$$

Sendo t_1, t_2, \dots, t_n , os n tempos de falhas ordenados por ordem crescente, onde t_i é o tempo de falha de ordem i observado.

Inicialmente, Waloddi Weibull usou a equação 4.5. É por este motivo que esta equação também é conhecida por “estimador de Weibull”.

O valor exato da mediana com diferentes combinações de i e n pode ser obtido usando métodos numéricos.

Aproximações simples da mediana têm sido propostas, entre os quais o estimador de Bernard que tem sido amplamente utilizado nos dias de hoje (O'Connor, 2012), (Tobias, 2011). O estimador de Bernard é dado pela seguinte equação:

$$\hat{F}_{(i)} = \frac{i-0,3}{n+0,4} \quad 4.6$$

Outra aproximação da mediana habitualmente utilizada é o estimador de Hazen, dada pela equação seguinte,

$$\hat{F}_{(i)} = \frac{i-0,5}{n} \quad 4.7$$

Muitos investigadores compararam vários estimadores, incluindo a média e os estimadores de Bernard e de Hazen para dados completos e com diferentes quantidades de dados (Fothergill, 1990), (Hossain, 2003), (Wang, 2001).

A maioria dos resultados obtidos permitiu verificar que o estimador de Bernard apresenta um menor enviesamento para diferentes quantidades de dados. Este tipo de resultado conduziu a que este estimador seja um dos mais utilizados atualmente.

Ao longo da última década outros estimadores foram apresentados; contudo não receberam a mesma atenção que os estimadores referidos anteriormente (Kirtay, 2012), (Yahaya, 2012).

b) Estimação de $F(t)$ para dados censurados

Conforme referido no capítulo 2.2.2 em muitos casos os dados obtidos contêm incertezas, ou seja, não é conhecido o momento exato em que ocorreu a falha. Os dados que contêm essa incerteza relativamente ao momento em que ocorreu o evento são considerados incompletos. Os dados incompletos podem ser classificados em censurados ou truncados.

Como fazer uso das informações fornecidas de uma amostra com dados censurados é o problema fundamental no processo de estimação e que irá afetar significativamente os resultados da estimação dos parâmetros. Obviamente, ignorar os dados censurados ou tratá-los como falhas fará com que o resultado não seja confiável, pois a informação obtida pelos dados censurados será perdida ou mal utilizada.

A influência de dados censurados deve ser refletida na estimativa da probabilidade de falha. Os estimadores mencionados anteriormente não devem ser utilizados diretamente.

O estimador de Kaplan-Meier é o mais antigo estimador não paramétrico aplicado a dados censurados (McCool, 2012).

Para a maioria do conjunto de dados, o método de Kaplan-Meier funciona muito bem, mas para o caso particular em que o último dado é uma falha, os resultados são menos satisfatórios.

O método de Kaplan-Meier para dados censurados é dado pela seguinte expressão:

$$\begin{cases} \hat{R}_{f,j} = \left(\frac{n - I_j}{n + 1 - I_j} \right) \hat{R}_{f,(j-1)} \\ \hat{F}_{f,j} = 1 - \hat{R}_{f,j} \end{cases} \quad 4.8$$

Em que $\hat{F}_{f,j}$ representa o valor estimado da função de distribuição acumulada para a j^{a} falha com dados censurados, I_j indica o número do evento correspondente à j^{a} falha e $\hat{R}_0 = 1$. A ocorrência de uma falha ou de um tempo censurado são ambos considerados como um evento. $\hat{R}_{f,j}$ é o complementar de $\hat{F}_{f,j}$.

Considerando uma amostra de tamanho n em que r falhas ($0 < r < n$) e $(n - r)$ dados censurados estão misturados ao longo do eixo do tempo. Sendo $t_1, t_2, \dots, t_i, \dots, t_n$ as observações ordenadas. t_i é considerado um evento e pode ser efetivamente uma falha ou um dado censurado. Em que $t_{f,1}, t_{f,2}, \dots, t_{f,j}, \dots, t_{f,r}$ ($1 \leq r \leq n$) são os tempos de falhas ordenados. Pela definição de I_j verifica-se que $t_{f,j} = t_{(I_j)}$.

De modo a ilustrar o cálculo com o método de Kaplan-Meier, apresenta-se um exemplo com cinco dados em que o quarto dado é censurado.

Os resultados obtidos estão apresentados na tabela 4.1.

Tabela 4.1 - Valor estimado da função probabilidade acumulada pelo método de Kaplan-Meier.

| j | Censurado | I_j | $R_{f,j}$ | $F_{f,j}$ |
|-----|-----------|-------|-----------|-----------|
| 1 | 0 | 1 | 0.80 | 0.20 |
| 2 | 0 | 2 | 0.60 | 0.40 |
| 3 | 0 | 3 | 0.40 | 0.60 |
| | 1 | | | |
| 4 | 0 | 5 | 0.0 | 1.00 |

Verifica-se que o valor estimado da função de probabilidade acumulada para o último dado é 1. Este resultado pressupõe que toda a população irá falhar até ao momento do último dado.

Da equação 4.8, verifica-se que se o último dado da amostra é uma falha, obtém-se $I_j = n$ e portanto, a função de probabilidade acumulada de falha é sempre igual a 1 para este ponto.

Não é razoável esperar que o último tempo de falha da amostra coincida com o tempo máximo de falha da população. Este método subestima o tempo de falha no extremo da distribuição.

Em 1960 Herd propôs um método para estimar o valor da função probabilidade acumulada de falha com dados censurados e em 1964 Johnson decompôs o método desenvolvido por Herd em dois passos. A combinação desses dois trabalhos é normalmente designada pelo método de Herd-Johnson (Zhang, 2006).

Este método na última década tem ganho popularidade para estimar a probabilidade de falha com dados censurados e é dado por:

$$\begin{cases} \hat{R}_{f,j} = \left(\frac{n+1-l_j}{n+2-l_j} \right) \hat{R}_{f,(j-1)} \\ \hat{F}_{f,j} = 1 - \hat{R}_{f,j} \end{cases} \quad 4.9$$

Na tabela 4.2 são apresentados os resultados obtidos pelo método de Herd-Johnson com os mesmos dados da tabela 4.1.

Tabela 4.2 - Valor estimado da função probabilidade acumulada pelo método de Herd-Johnson.

| j | <i>Censurado</i> | l_j | $R_{f,j}$ | $F_{f,j}$ |
|-----|------------------|-------|-----------|-----------|
| 1 | 0 | 1 | 0.83 | 0.17 |
| 2 | 0 | 2 | 0.67 | 0.33 |
| 3 | 0 | 3 | 0.50 | 0.50 |
| | 1 | | | |
| 4 | 0 | 5 | 0.25 | 0.75 |

Verifica-se que o problema que surgiu com o método de Kaplan-Meier não surge com este método.

Em 1972 Nelson apresentou um método para estimar o valor da função probabilidade acumulada de falha com dados censurados, que ficou designado de estimador de Nelson (McCool, 2012), Nelson (1982) e é dado por:

$$\begin{cases} \hat{R}_{f,j} = \exp\left(-\frac{1}{n+1-l_j}\right) \hat{R}_{f,(j-1)} \\ \hat{F}_{f,j} = 1 - \hat{R}_{f,j} \end{cases} \quad 4.10$$

Na tabela 4.3 são apresentados os resultados obtidos pelo método de Nelson com os mesmos dados da tabela 4.1.

Tabela 4.3 - Valor estimado da função probabilidade acumulada pelo método de Nelson.

| j | <i>Censurado</i> | l_j | $R_{f,j}$ | $F_{f,j}$ |
|-----|------------------|-------|-----------|-----------|
| 1 | 0 | 1 | 0.82 | 0.18 |
| 2 | 0 | 2 | 0.64 | 0.36 |
| 3 | 0 | 3 | 0.46 | 0.54 |
| | 1 | | | |
| 4 | 0 | 5 | 0.17 | 0.83 |

4.2.2 - Métodos analíticos de estimação

Existe uma grande variedade de métodos analíticos para estimar os parâmetros de Weibull. Os principais métodos são: dos mínimos quadrados e de máxima verosimilhança (Hossain, 2003), (Montanari, 1998), (Olteanu, 2010), (Procaccia, 2011). Existem outros métodos, como por exemplo, o método dos momentos, porém não é abordado neste trabalho (O'Connor, 2012), (Procaccia, 2011).

4.2.2.1 - Método de estimação de mínimos quadrados

O método de estimação de mínimos quadrados é basicamente a versão analítica do método gráfico de probabilidade de Weibull. Na prática estes dois métodos são normalmente utilizados juntos (Abernethy, 2006), (Montgomery, 2003), (O'Connor, 2012), (Seber, 2003).

Ao combinar os dois métodos está-se basicamente a usar a técnica da regressão por mínimos quadrados para gerar a linha reta que melhor se ajusta aos pontos colocados no gráfico de probabilidade de Weibull em vez de traçar a “olho”.

As vantagens de combinar os dois métodos são óbvias:

- elimina a subjetividade de traçar a linha que melhor se ajusta aos pontos a “olho”, de modo a melhorar a eficiência da estimação;
- fornece ao método de estimação de mínimos quadrados uma apresentação gráfica que pode servir como modelo de validação e identificação de pressupostos errados, além da estimação dos parâmetros de Weibull.

Este método utiliza a regressão de mínimos quadrados para estimar os dois parâmetros da equação 4.2.

Assim, a estimativa do valor de η e β pode ser obtida pela estimativa dos coeficientes do modelo de regressão linear simples na forma $Y = AX + B + e$, onde e é o termo do erro.

Para uma amostra com dados completos $t_1, t_2, \dots, t_i, \dots, t_n$, os valores de X e Y podem ser

4 - Estimação dos parâmetros da distribuição de Weibull

obtidos por,

$$x_i = \ln(t_i) \quad 4.11$$

$$y_i = \ln[-\ln(1 - \hat{F}_i)] \quad 4.12$$

Para uma amostra com dados censurados onde $t_{f,1}, t_{f,2}, \dots, t_{f,j}, \dots, t_{f,r}$ correspondem aos tempos de falhas, o valor de X e Y pode ser obtido por,

$$x_i = \ln(t_{f,j}) \quad 4.13$$

$$y_i = \ln[-\ln(1 - \hat{F}_{f,j})] \quad 4.14$$

Conforme se verifica nas equações anteriores, o método de estimação de mínimos quadrados precisa, do mesmo modo do que o método gráfico de probabilidade de Weibull, da estimação do valor da função de probabilidade acumulada de falha, $F(t)$, para cada tempo de falha.

Os métodos utilizados para obter os valores de \hat{F}_i e $\hat{F}_{f,j}$ foram descritos na Seção 4.2.1.1.

O método de estimação por mínimos quadrados consiste em minimizar o quadrado das diferenças entre os valores observados de uma amostra e os seus respetivos valores esperados de forma a maximizar o grau de ajuste aos dados obtidos, ou seja,

$$\min S = \sum_{i=1}^r [y_i - (Ax_i + B)]^2 \quad 4.15$$

para dados completos, $r = n$.

Ao efetuar as derivadas parciais de S em ordem a A e B , respetivamente e igualando a zero obtém-se o seguinte resultado,

$$\begin{cases} \hat{A} = \frac{\sum_{i=1}^r [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^r (x_i - \bar{x})^2} = \frac{r \sum_{i=1}^r (x_i y_i) - \sum_{i=1}^r x_i \cdot \sum_{i=1}^r y_i}{r \sum_{i=1}^r x_i^2 - \sum_{i=1}^r x_i} \\ \hat{B} = \bar{y} - \hat{A} \bar{x} = \frac{\sum_{i=1}^r y_i - \hat{A} \sum_{i=1}^r x_i}{r} \end{cases} \quad 4.16$$

Onde,

$$\bar{x} = \sum_{i=1}^r \frac{x_i}{r}, \text{ é a média de } x$$

e

$$\bar{y} = \sum_{i=1}^r \frac{y_i}{r}, \text{ é a média de } y.$$

Dado que $A = \beta$ e $B = -\beta \ln(\eta)$, o valor estimado de η e β é obtido pelas seguintes equações,

$$\begin{cases} \hat{\beta} = \frac{\sum_{i=1}^r [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^r (x_i - \bar{x})^2} = \frac{r \sum_{i=1}^r (x_i y_i) - \sum_{i=1}^r x_i \cdot \sum_{i=1}^r y_i}{r \sum_{i=1}^r x_i^2 - \sum_{i=1}^r x_i} \\ \hat{\eta} = \exp\left(-\frac{\bar{y} - \hat{\beta} \bar{x}}{\hat{\beta}}\right) = \exp\left(-\frac{\sum_{i=1}^r y_i - \hat{\beta} \sum_{i=1}^r x_i}{r \hat{\beta}}\right) \end{cases} \quad 4.17$$

A equação 4.17 pode ser aplicada para dados completos e para dados censurados. Para dados completos $r = n$.

O método de estimação por mínimos quadrados é considerado pela opinião tradicional entre os investigadores, como um método simples e impreciso (similar ao método de estimação gráfica) e é habitualmente indicado para fornecer os valores iniciais dos parâmetros para os outros métodos de estimação mais sofisticados, como por exemplo o método de máxima verosimilhança (Procaccia, 2011), (Rinne, 2009).

4.2.2.1.1 - Simulação de Monte Carlo

Nesta secção, a simulação pelo método de Monte Carlo realizou-se com o objetivo de encontrar o melhor método para obter o valor de $F(t)$, com dados completos e censurados, de entre os métodos descritos na secção 4.2.1.1.

A simulação de Monte Carlo permitiu simular o comportamento de cada um dos métodos para diferentes tamanhos da amostra e percentagem de dados censurados e assim avaliar e comparar qual dos métodos é mais adequado de utilizar para cada um dos casos, conforme referido no anexo A.

Como foi referido anteriormente na Seção 2.2.4, o enviesamento, a variância e o erro quadrático médio (EQM) dos estimadores são os critérios mais comuns para avaliar o desempenho de um método de estimação. Com a simulação pelo método de Monte Carlo, o enviesamento, a variância e o erro quadrático médio dos estimadores da distribuição podem ser determinados e assim comparar os diferentes métodos de estimação.

Os resultados da simulação são apresentados sob diferentes combinações dos parâmetros experimentais. No anexo A é explicado a seleção de cada um dos parâmetros.

a) Procedimento da simulação de Monte Carlo para dados completos

O procedimento passo-a-passo da simulação para dados completos é descrito como se segue:

Passo 1: Gerar n números aleatórios t_1, t_2, \dots, t_n a partir de uma distribuição de Weibull, dados os parâmetros η_T e β_T .

Passo 2: Calcular o valor de y_i para cada método (média, Bernard e Hazen).

Passo 3: Com os resultados gerados no passo anterior, estimar η e β , pela utilização do método de estimação de mínimos quadrados (capítulo 4.2.2.1).

Passo 4: Repetir o passo 1 ao passo 3, M vezes (M é designado por número de iterações ou número de repetições).

Passo 5: Calcular a função pivotal (equação A.1 e A.2). Calcular o enviesamento, a variância e o erro quadrático médio do estimador com as fórmulas seguintes, respetivamente (Zhang, 2007).

$$B(\hat{\theta}) = \bar{\theta} - \theta$$

$$Var(\hat{\theta}) = \frac{1}{M} \sum_{i=1}^M (\hat{\theta}_i - \bar{\theta})^2 \quad 4.18$$

$$EQM(\hat{\theta}) = \frac{1}{M} \sum_{i=1}^M (\hat{\theta}_i - \theta)^2$$

Onde θ pode ser substituído por η e β e $\bar{\theta} = \frac{1}{M} \sum_{i=1}^M \hat{\theta}_i$.

b) Procedimento da simulação de Monte Carlo para dados censurados

O procedimento passo-a-passo da simulação para dados censurados é descrito como se segue:

Passo 1: Gerar n números aleatórios t_1, t_2, \dots, t_n a partir de uma distribuição de Weibull, dados os parâmetros η_T e β_T .

Passo 2: A partir da amostra completa t_1, t_2, \dots, t_n gerada no passo anterior, selecionar aleatoriamente $n-r$ tempos censurados, denominados por $t_{c,k}$ ($k = 1, 2, \dots, n-r$). As restantes observações são tempos de falha, denominados por $t_{f,j}$ ($j = 1, 2, \dots, r$).

Passo 3: Misturar aleatoriamente os tempos de falha com os tempos censurados para produzir a amostra censurada.

Passo 4: Calcular o valor de y_i para cada método (Kaplan-Meier, Herd-Johnson e Nelson).

Passo 5: Para a amostra atual, estimar η e β , pela utilização do método de estimação de mínimos quadrados (capítulo 4.2.2.1).

Passo 6: Repetir o passo 1 ao passo 7, M vezes (M é designado por número de iterações ou número de repetições).

Passo 7: Calcular a função pivotal (equação A.1 e A.2). Calcular o enviesamento, a variância e o erro quadrático médio do estimador com as fórmulas 4.18.

Resultado da simulação do método dos mínimos quadrados

a) Resultados da simulação com dados completos

Foram examinados, pela simulação de Monte Carlo, três estimadores para obter o valor de $F(t)$, nomeadamente a média (equação 4.5), o estimador de Bernard (equação 4.6) e o estimador de Hazen (equação 4.7). As comparações entre os três métodos focaram-se em amostras de pequeno e médio tamanho dado que o método dos mínimos quadrados habitualmente apresenta piores resultados nestas condições.

Na tabela 4.4 são apresentados os parâmetros da simulação para a estimação do parâmetro de forma, β , com dados completos, nomeadamente, os valores dos parâmetros reais de η e β (denominados de η_T e β_T), o tamanho da amostra n e o número de iteração M , conforme referido no anexo A.

Tabela 4.4 - Parâmetros da simulação para a estimação do parâmetro de forma, β , com dados completos.

| Parâmetros | Valores |
|------------|------------------------------|
| η_T | 1 |
| β_T | 1 |
| n | 4, 7, 10 - 100 (de 10 em 10) |
| M | 10000 |

Os resultados são apresentados através de figuras sendo um dos eixos a função pivotal, o que permite facilmente verificar o desempenho dos métodos.

O comportamento do enviesamento é idêntico à função pivotal. Assim, o resultado ideal é igual a 1, o que significa que não apresenta enviesamento. Quanto mais afastado de 1 o resultado estiver, maior é o enviesamento. Se o resultado da função pivotal for maior do que 1 significa que o valor estimado é maior do que o valor real. Se o resultado for menor do que 1 significa que o valor estimado é menor do que o valor real. Se o resultado for igual a 1 significa que o valor estimado é igual ao valor real.

Os resultados de $\hat{\beta}$ são apresentados na figura 4.1 através da função pivotal. O erro quadrático médio (EQM) dos estimadores não é apresentado apenas é comentado os seus resultados quando assim se justifica.

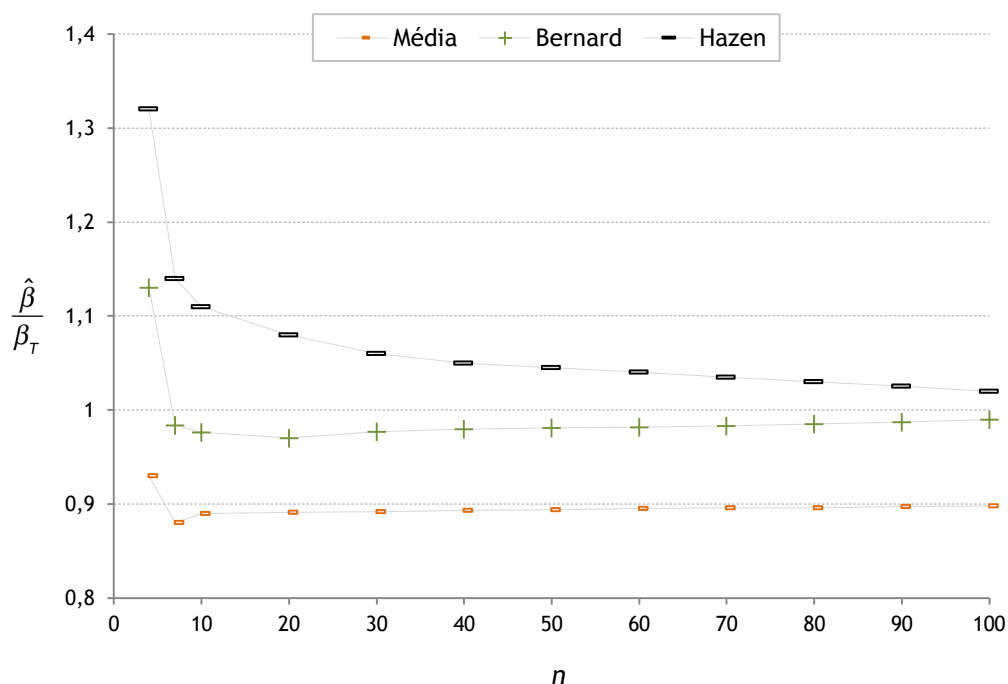


Figura 4.1 - Comparação da estimação do parâmetro de forma, β , para dados completos ($\eta_T = \beta_T = 1$).

Nenhum dos métodos tem o melhor desempenho para todos os tamanhos das amostras. Além disso, nenhum deles é não enviesado. Quando o tamanho da amostra é muito pequeno ($n = 4$) a média apresenta melhores resultados, no entanto, é o pior para $n \geq 7$. O estimador de Bernard tem o melhor desempenho para $n \geq 7$. O estimador de Hazen superestima $\hat{\beta}$. O estimador de Bernard subestima $\hat{\beta}$ quando $n \geq 7$ e apresenta um resultado quase igual a um quando é gerado para $n \geq 90$.

O comportamento da variância e do EQM é semelhante ao enviesamento, nenhum dos métodos apresenta os melhores resultados para todos os tamanhos das amostras. Para $n < 10$ o EQM de $\hat{\beta}$ gerado pelo estimador de Bernard é menor do que o resultado dos outros estimadores, especialmente para $n=4$. Para $n \geq 10$ os resultados do EQM de $\hat{\beta}$ gerado por todos os métodos são semelhantes.

Na tabela 4.5 são apresentados os parâmetros da simulação para a estimação do parâmetro de escala, η .

Tabela 4.5 - Parâmetros da simulação para a estimação do parâmetro de escala, η , com dados completos.

| Parâmetros | Valores |
|------------|----------------|
| η_T | 1 |
| β_T | 0.5, 1, 3 |
| n | 4, 7, 10 - 100 |
| M | 10000 |

Os resultados de $\hat{\eta}$ são apresentados da figura 4.2 à 4.4.

Comparando as três figuras (figuras 4.2, 4.3 e 4.4), verifica-se que o resultado de $\hat{\eta}$, em todos os estimadores, diminui com o aumento de β_T . A tendência de $\hat{\eta}$ em função de n varia com β_T para todos os estimadores. A tendência de $\hat{\eta}$ para o estimador de Hazen e Bernard para $\beta_T = 3$ é diferente da tendência para $\beta_T = 0.5$ e para $\beta_T = 1$. Para todos os estimadores, a estimação de $\hat{\eta}$ é consistente para $\beta_T = 0.5$, mas inconsistente para $\beta_T = 1$ e $\beta_T = 3$. Para $\beta_T = 0.5$ o estimador de Hazen supera os outros estimadores em todos os tamanhos da amostra e todos os estimadores superestimam $\hat{\eta}$.

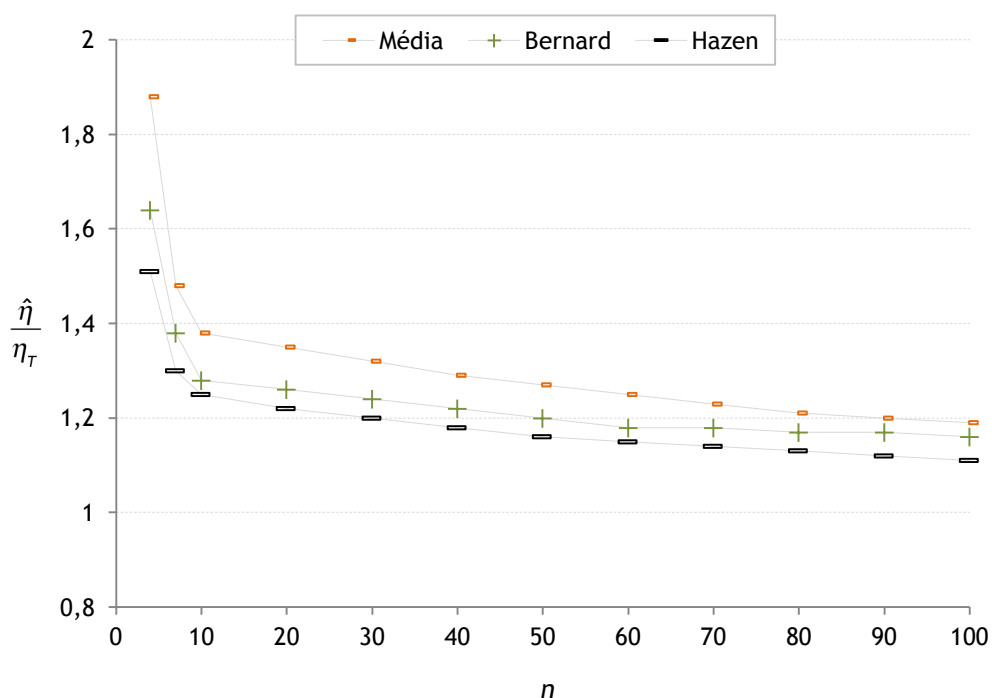


Figura 4.2 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = 1$, $\beta_T = 0.5$).

4 - Estimação dos parâmetros da distribuição de Weibull

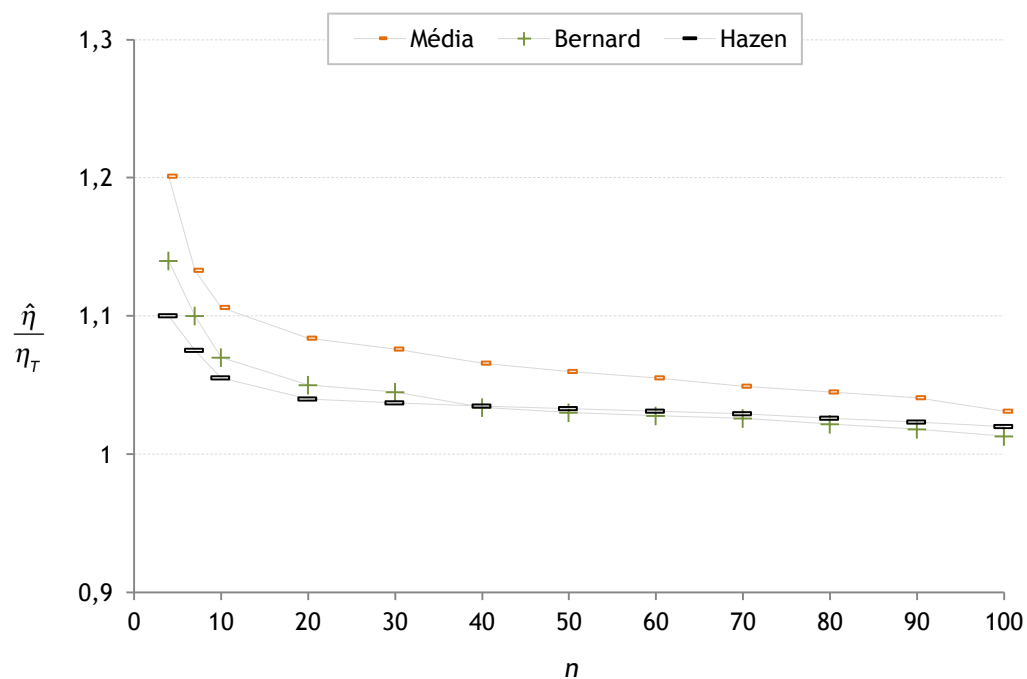


Figura 4.3 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = \beta_T = 1$).

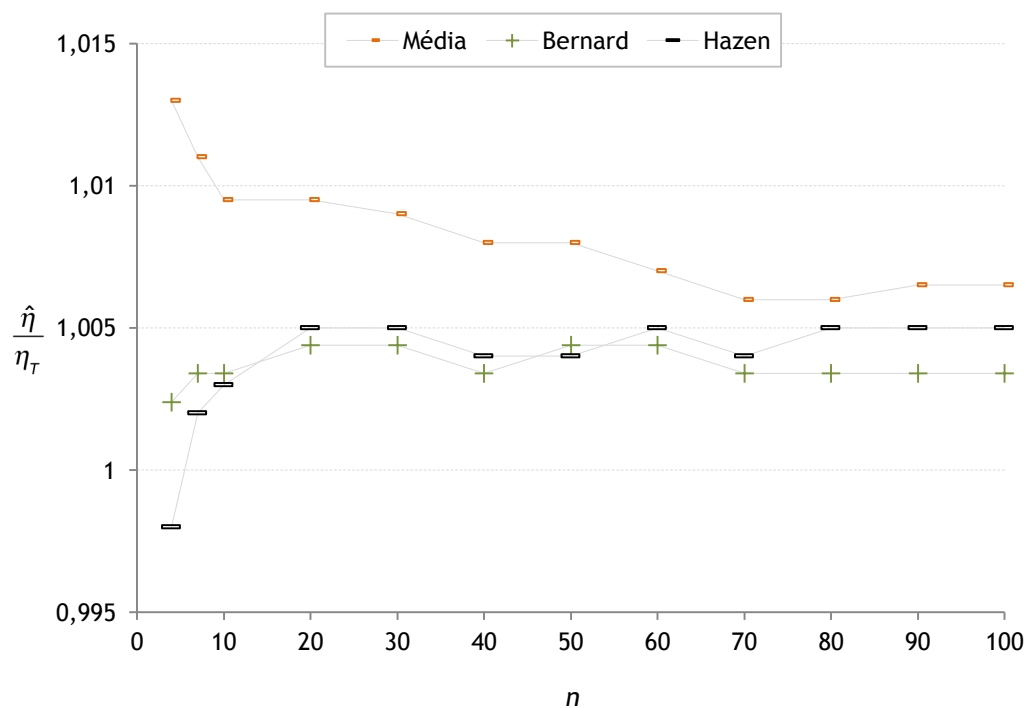


Figura 4.4 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = 1$, $\beta_T = 3$).

O enviesamento de $\hat{\eta}$, para todos os métodos, é maior que 10% para $\beta_T = 0.5$. Para $\beta_T = 1$ e $\beta_T = 3$ o estimador de Bernard apresenta o melhor desempenho, seguido pelo estimador de Hazen, a média apresenta o pior desempenho. Para $\beta_T = 3$ o estimador de Hazen subestima $\hat{\eta}$ para $n < 7$. Para $n \geq 20$, o estimador de Bernard é o melhor.

A variância e o EQM para $\hat{\eta}$ depende de β_T e de n . Para todos os métodos, o valor da variância e do EQM diminui à medida que β_T ou o tamanho da amostra, n aumenta. Para todos os valores de β_T , o valor do EQM de $\hat{\eta}$ para o estimador de Bernard é menor do que para os outros estimadores, especialmente para $\beta_T = 0,5$ e $n \leq 10$.

Considerando o enviesamento e o EQM em conjunto, o estimador de Bernard apresenta o melhor resultado na maioria dos casos.

b) Resultados da simulação com dados censurados

Foram examinados, pela simulação de Monte Carlo, três estimadores para obter o valor de $F(t)$, nomeadamente o estimador de Kaplan-Meier (equação 4.8), o estimador de Herd-Johnson (equação 4.9) e o estimador de Nelson (equação 4.10).

Na tabela 4.6 são apresentados os parâmetros da simulação para a estimação do parâmetro de forma, β , com dados censurados, nomeadamente, os valores dos parâmetros reais de η e β (denominados de η_T e β_T), o tamanho da amostra n , percentagem de dados censurados c e o número de iteração M , conforme referido no anexo A.

Tabela 4.6 - Parâmetros da simulação para a estimação do parâmetro de forma, β , com dados censurados.

| Parâmetros | Valores |
|------------|----------------|
| η_T | 1 |
| β_T | 1 |
| n | 4, 7, 10 - 100 |
| c | 20%, 60% |
| M | 10000 |

Os resultados de $\hat{\beta}$ são apresentados nas figuras 4.5 e 4.6, com um nível de censura, $c=20\%$ e $c=60\%$, respetivamente, através da função pivotal.

O comportamento do enviesamento é idêntico à função pivotal. O erro quadrático médio (EQM) dos estimadores não é apresentado, apenas é comentado os seus resultados quando assim se justifica.

4 - Estimação dos parâmetros da distribuição de Weibull

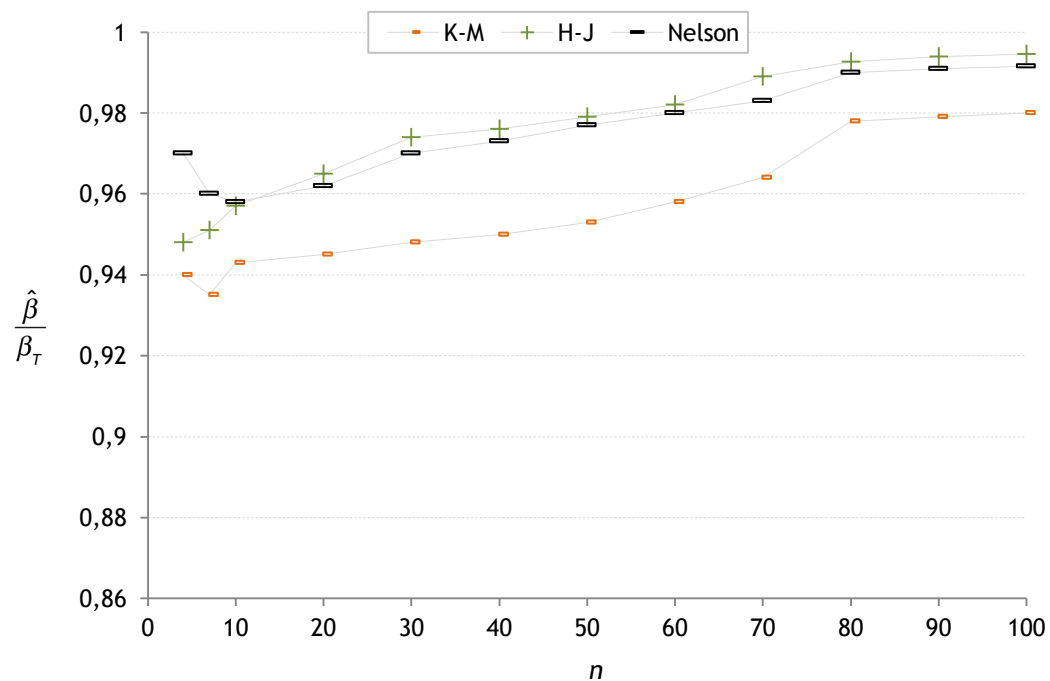


Figura 4.5 - Comparação da estimação do parâmetro de forma, β , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 20\%$.

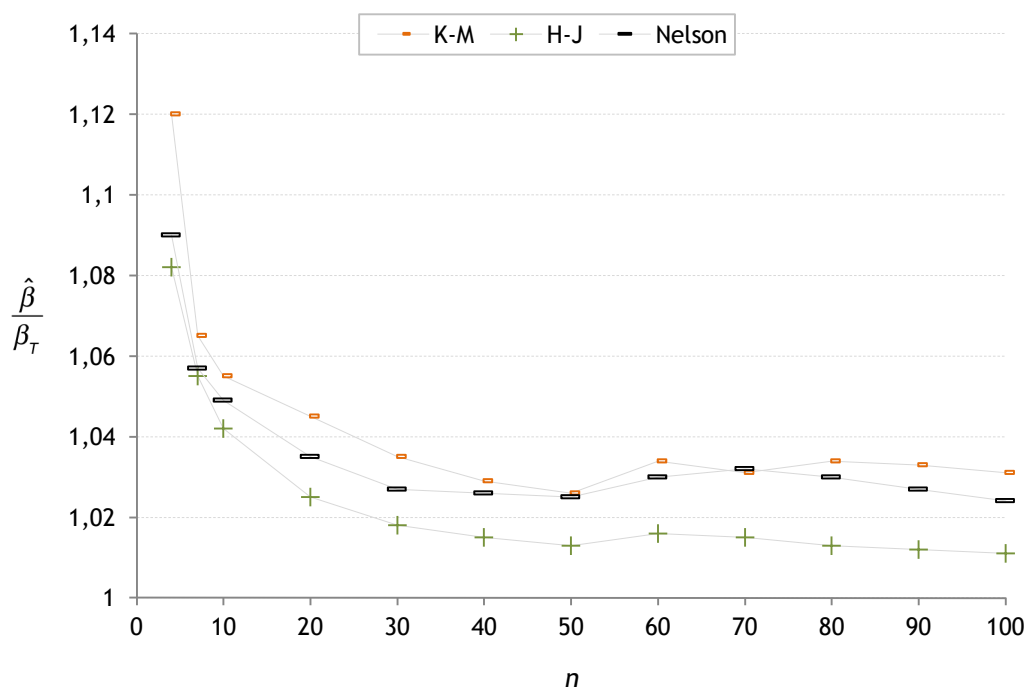


Figura 4.6 - Comparação da estimação do parâmetro de forma, β , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 60\%$.

Para as amostras com dados censurados as propriedades do parâmetro $\hat{\beta} / \beta_T$ continuam a ser válidas. Portanto, este parâmetro pode ser utilizado para verificar o desempenho da simulação, como se fez para os dados completos.

Os resultados das amostras com níveis de censura baixo (20%) e das amostras com níveis de censura elevada (60%) são bastantes diferentes. Nenhum dos estimadores supera os outros em todas as combinações experimentais, relativamente ao enviesamento e ao EQM de $\hat{\beta}$.

Os estimadores Herd-Johnson e Nelson apresentam comportamentos semelhantes para $c=20\%$. Quando o nível de censura é elevado (60%) o enviesamento do estimador Herd-Johnson é inferior comparativamente aos outros estimadores.

Relativamente ao EQM, o estimador Herd-Johnson apresenta os melhores resultados na maioria das condições, exceto quando $c=20\%$ e $n \geq 20$.

Na tabela 4.7 são apresentados os parâmetros da simulação para a estimação do parâmetro de forma, η , com dados censurados.

Tabela 4.7 - Parâmetros da simulação para a estimação do parâmetro de forma, η , com dados censurados.

| Parâmetros | Valores |
|------------|----------------|
| η_T | 1 |
| β_T | 1, 3 |
| n | 4, 7, 10 - 100 |
| c | 20%, 60% |
| M | 10000 |

Os resultados de $\hat{\eta}$ são apresentados nas figuras 4.7 e 4.8, com um nível de censura, $c=20\%$ e $c=60\%$, respetivamente, para $\beta_T=1$ e nas figuras 4.9 e 4.10, com um nível de censura, $c=20\%$ e $c=60\%$, respetivamente, para $\beta_T=3$, através da função pivotal.

O enviesamento de $\hat{\eta}$ é grande para $c=60\%$. O enviesamento pelo estimador Herd-Johnson apresenta na maioria das situações o melhor desempenho, exceto para $n \geq 80$ ($\eta_T=1$, $\beta_T=3$, $c=60\%$).

O comportamento dos resultados do desvio padrão e do MSE para $\hat{\eta}$ é semelhante ao obtido para o enviesamento. O estimador de Nelson tem melhores resultados quando o nível de censura é baixo e o estimador de Herd-Johnson tem melhores resultados do que o estimador de Nelson quando o nível de censura é elevado.

4 - Estimação dos parâmetros da distribuição de Weibull

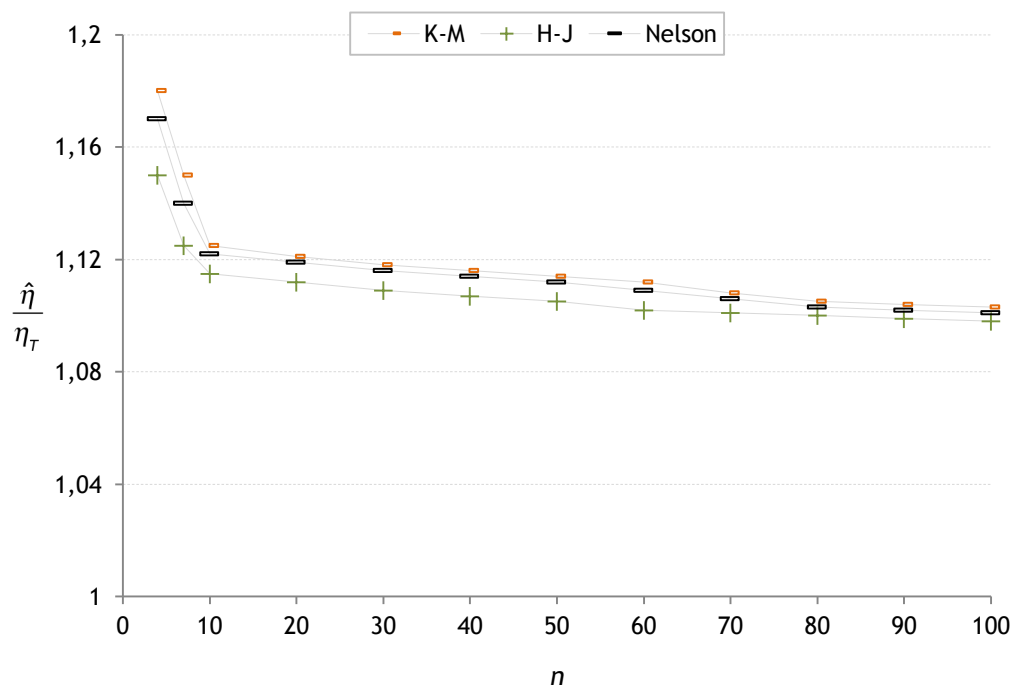


Figura 4.7 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 20\%$.

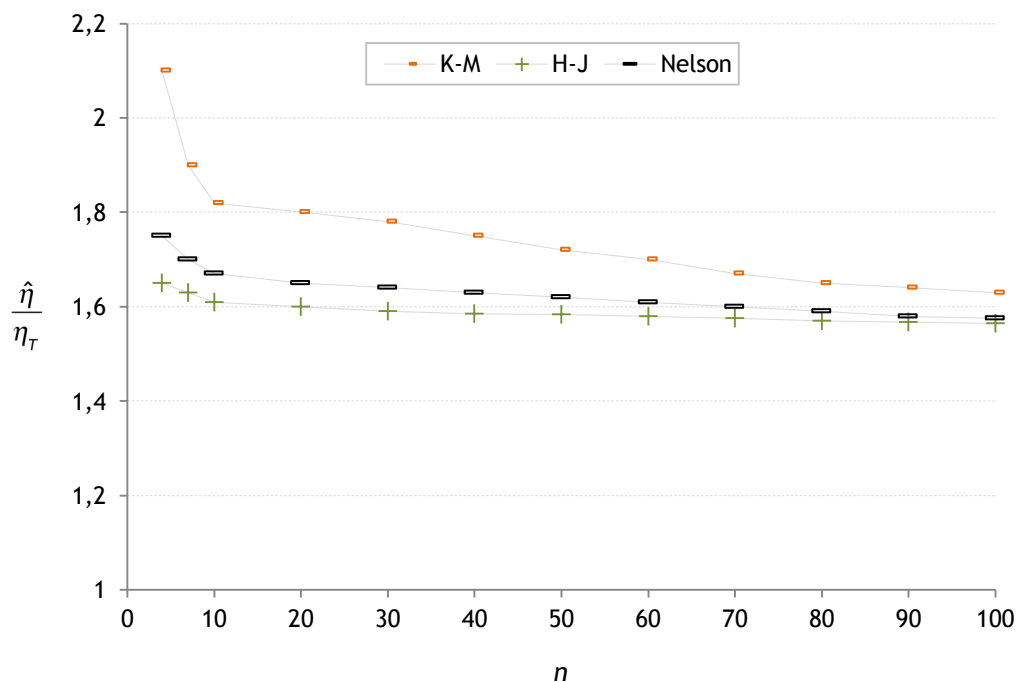


Figura 4.8 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 60\%$.

4 - Estimação dos parâmetros da distribuição de Weibull

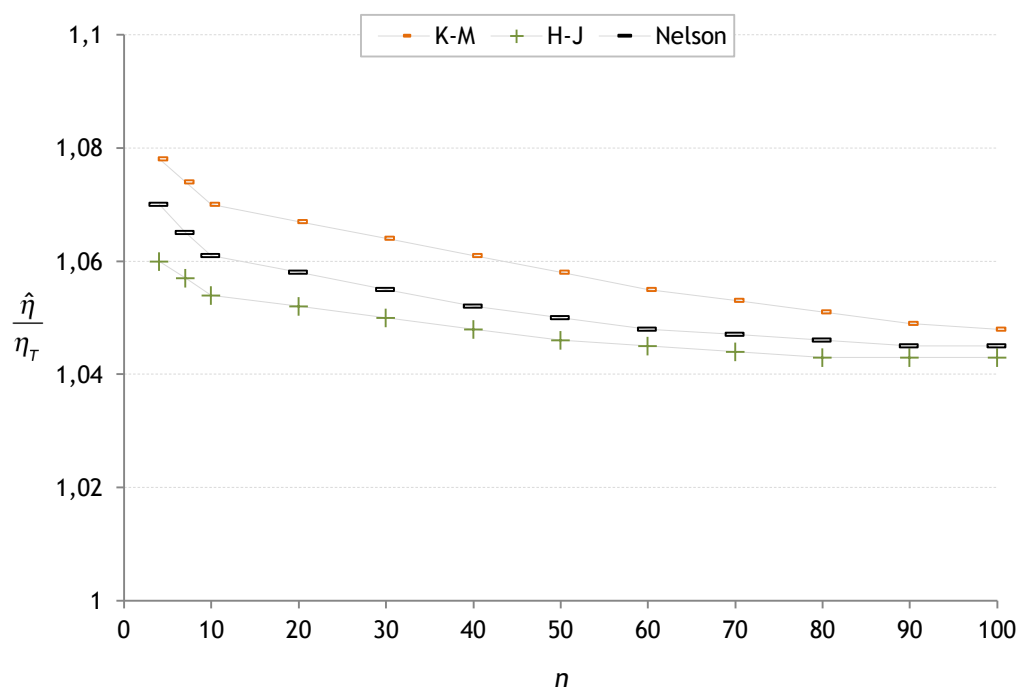


Figura 4.9 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = 1$, $\beta_T = 3$), com $c = 20\%$.

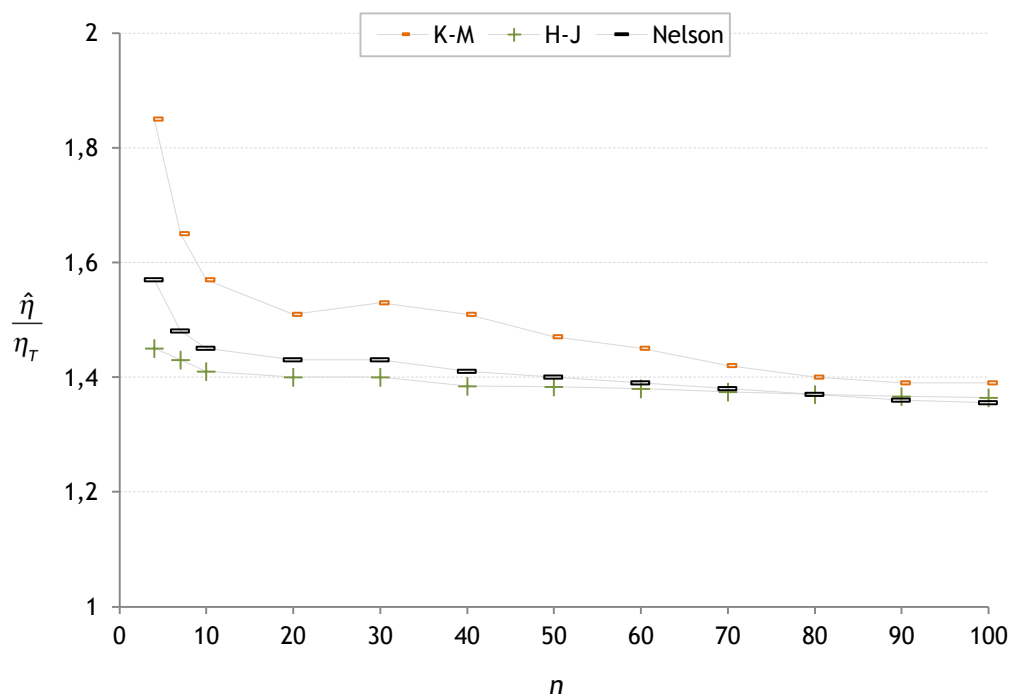


Figura 4.10 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = 1$, $\beta_T = 3$), com $c = 60\%$.

Considerando o enviesamento e o EQM em conjunto, o estimador de Herd-Johnson apresenta o melhor resultado na maioria dos casos.

4.2.2.2 - Método de estimação de máxima verosimilhança

O método de máxima verosimilhança é atualmente o método mais popular de estimação (Held, 2014).

Este método geralmente é creditado a R. A. Fisher (1890 - 1962), embora as suas raízes remontem a J. Y. Lambert (1760), J. L. Lagrange (1770) e Daniel Bernoulli (1778) no século XVIII. Contudo foi Fisher, em 1922, quem publicou este método na forma que hoje se conhece, como uma alternativa para o método dos momentos e para o método dos mínimos quadrados (descrito anteriormente) (Edwards, 1974).

O método de máxima verosimilhança é considerado por ser um dos métodos mais versáteis e fiáveis (Chambers, 2012), (Held, 2014). As propriedades deste método são apresentadas neste sub-capítulo.

O método de estimação de máxima verosimilhança permite estimar os parâmetros desconhecidos de um modelo estatístico. Estes parâmetros são obtidos através da maximização da função de verosimilhança do modelo em análise.

Seja $x_i = x_1, x_2, \dots, x_n$ uma amostra de n observações independentes da variável aleatória X , de uma distribuição cuja função densidade de probabilidade, $f(x, \theta)$, é definida pela equação 3.2, onde $\theta = \theta_1, \theta_2, \dots, \theta_k$ é o vector dos parâmetros desconhecidos.

Uma amostra aleatória é constituída por n observações de uma mesma população. Se $f(x)$ for a função densidade de probabilidade da subjacente população, então $f(x_i)$ é a função densidade de probabilidade do $i^{\text{ésimo}}$ valor. Dado que a amostra compreende n valores independentes, a distribuição de probabilidade conjunta da amostra será o produto das n idênticas e independentes distribuições de probabilidade (Ebeling, 1997), (Rohde, 2014), ou seja,

$$f_{x_1, x_2, \dots, x_n}(x_1, x_2, \dots, x_n) = f(x_1) \cdot f(x_2) \dots f(x_n) \quad 4.19$$

A função de verosimilhança é a função densidade de probabilidade da distribuição conjunta da variável aleatória X . Assim, a função de verosimilhança de θ , para dados completos é definida por,

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n f(\theta_1, \theta_2, \dots, \theta_k | x_i) \quad 4.20$$

O objetivo do método de estimação de máxima verosimilhança é determinar o vector dos parâmetros desconhecidos θ , que maximiza a equação 4.20, ou seja,

$$\max L(\theta_1, \theta_2, \dots, \theta_k) = \max \prod_{i=1}^n f(\theta | x_i) \quad 4.21$$

Em muitas situaes  mais fcil obter a maximizao do logaritmo da funo de verosimilhana e dado que a funo logaritmo  uma funo montona crescente,  equivalente maximizar a funo de verosimilhana ou a funo de log-verosimilhana dada por,

$$\begin{aligned} \max \ln L(\theta_1, \theta_2, \dots, \theta_k) &= \max \ln \left[\prod_{i=1}^n f(\theta | x_i) \right] \\ &= \max \sum_{i=1}^n [\ln(f(\theta | x_i))] \end{aligned} \quad 4.22$$

Considerando que a funo de verosimilhana  diferencivel e satisfaa as condies de regularidade, a estimao dos parmetros da distribuico pode ser obtida pela derivada parcial do logaritmo da funo de verosimilhana igual a zero, conforme indicado na seguinte equao,

$$\frac{\partial \ln L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_i} = 0; \quad i = 1, 2, \dots, k \quad 4.23$$

 necessrio verificar se a segunda derivada  negativa para garantir que os resultados obtidos de 4.23 correspondem a um ponto mximo, ou seja,

$$\left. \frac{\partial^2 \ln L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_i^2} \right|_{\theta=\hat{\theta}} < 0; \quad i = 1, 2, \dots, k \quad 4.24$$

$$\left. \frac{\partial^2 \ln L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_l \partial \theta_j} \right|_{\theta=\hat{\theta}} < 0; \quad l = 1, 2, \dots, k; \quad j = 1, 2, \dots, k, \text{ para todo } l \neq j \quad 4.25$$

Normalmente as solues obtidas so mximos, mas nem sempre o valor obtido por derivao constitui um mximo global. Assim  necessria a sua verificao.

Em muitas situaes prticas a funo de verosimilhana est associada a modelos complexos e a equao de verosimilhana no apresenta soluo analtica explcita, em que s  possvel a sua resoluo atravs de mtodos numricos.

Os tempos de falha podem ser representados por $t_i = t_1, t_2, \dots, t_n$. Considera-se que cada falha representa um dado independente de uma mesma populao representativa da distribuico de Weibull com o parmetro de escala, η e o parmetro de forma, β . A funo de verosimilhana para a distribuico de Weibull com dados completos  obtida pela substituico da equao 3.50 na equao 4.20, conforme indicado na equao seguinte,

4 - Estimação dos parâmetros da distribuição de Weibull

$$L(\eta, \beta) = \prod_{i=1}^n \left(\frac{\beta}{\eta} \left(\frac{t_i}{\eta} \right)^{\beta-1} \exp \left(- \left(\frac{t_i}{\eta} \right)^{\beta} \right) \right) \quad 4.26$$

$$= \left(\frac{\beta}{\eta^{\beta}} \right)^n \prod_{i=1}^n \left(t_i^{\beta-1} \exp \left(- \left(\frac{t_i}{\eta} \right)^{\beta} \right) \right) \quad 4.27$$

Ao logaritmizar a equação anterior vem,

$$\ln L(\eta, \beta) = l(\eta, \beta) = n \ln \beta - n \beta \ln \eta + (\beta - 1) \sum_{i=1}^n (\ln t_i) - \sum_{i=1}^n \left(\frac{t_i}{\eta} \right)^{\beta} \quad 4.28$$

De modo a determinar os pontos máximos da equação anterior e assim obter a estimação dos parâmetros η e β da distribuição, é necessário resolver as derivadas parciais do logaritmo da função de verosimilhança e igualar a zero, conforme indicado nas seguintes equações,

$$\frac{\partial l(\eta, \beta)}{\partial \eta} = -\frac{n\beta}{\eta} + \frac{\beta}{\eta} \sum_{i=1}^n \left(\frac{t_i}{\eta} \right)^{\beta} = 0 \quad 4.29$$

$$\frac{\partial l(\eta, \beta)}{\partial \beta} = \frac{n}{\beta} - n \ln \eta + \sum_{i=1}^n \ln t_i - \sum_{i=1}^n \left[\left(\frac{t_i}{\eta} \right)^{\beta} \ln \left(\frac{t_i}{\eta} \right) \right] = 0 \quad 4.30$$

Assim a equação 4.29 pode ser apresentada pela seguinte expressão,

$$\hat{\eta} = \left[\frac{1}{n} \sum_{i=1}^n t_i^{\hat{\beta}} \right]^{\frac{1}{\hat{\beta}}} \quad 4.31$$

e a equação 4.30 pela seguinte equação e resolvida numericamente,

$$\frac{1}{\beta} + \frac{1}{n} \sum_{i=1}^n \ln t_i - \frac{\sum_{i=1}^n t_i^{\beta} \ln t_i}{\sum_{i=1}^n t_i^{\beta}} = 0 \quad 4.32$$

A segunda derivada deverá ser negativa para garantir que os resultados obtidos correspondam a um ponto máximo. As equações da segunda derivada são as seguintes,

$$\frac{\partial^2 l(\eta, \beta)}{\partial \eta^2} = \frac{\beta}{\eta^2} \left[n + (\beta - 1) \sum_{i=1}^n \left(\frac{t_i}{\eta} \right)^{\beta} \right] \quad 4.33$$

$$\frac{\partial^2 l(\eta, \beta)}{\partial \eta \partial \beta} = \frac{\partial^2 l(\eta, \beta)}{\partial \beta \partial \eta} = -\frac{1}{\eta} \left[n - \sum_{i=1}^n \left(\frac{t_i}{\eta} \right)^\beta - \beta \sum_{i=1}^n \left(\left(\frac{t_i}{\eta} \right)^\beta \ln \left(\frac{t_i}{\eta} \right) \right) \right] \quad 4.34$$

$$\frac{\partial^2 l(\eta, \beta)}{\partial \beta^2} = -\frac{n}{\beta^2} - \sum_{i=1}^n \left[\left(\frac{t_i}{\eta} \right)^\beta \left(\ln \left(\frac{t_i}{\eta} \right) \right)^2 \right] \quad 4.35$$

a) Método de estimação de máxima verosimilhança para dados censurados à direita tipo 1

Conforme referido no capítulo 2, em muitas situações práticas os dados contêm informações incompletas. Nos dados censurados à direita o tempo de falha das unidades com dados censurados apenas se sabe que é superior ao tempo de funcionamento correspondente à conclusão do registo da informação. No caso particular dos dados censurados à direita classificados em censura tipo 1, o registo da informação é interrompido a um tempo pré-determinado $Cd > 0$, tal que, t_i é observado se ocorrer antes de Cd , caso contrário, apenas se sabe que o tempo de falha é maior que o tempo de observação.

Seja $t_i = t_1, t_2, \dots, t_n$, em que r registos correspondem a tempos de falha e $(n - r)$ não correspondem a tempos de falha dentro do tempo limite t .

Assim, a caracterização dos dados observados são definidos pela variável δ_i , onde,

$$\delta_i = \begin{cases} 1, & \text{para dados não censurados} \\ 0, & \text{para dados censurados à direita} \end{cases}$$

Seja a função densidade de probabilidade, $f(x, \theta)$ e a função de probabilidade acumulada, $F(x, \theta)$, a função de verosimilhança para dados censurados à direita tipo 1 é dada por (Guure, 2013),

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{\delta_i=1} f(\theta_1, \theta_2, \dots, \theta_k | x_i) \prod_{\delta_i=0} [1 - F(\theta_1, \theta_2, \dots, \theta_k | x_i)] \quad 4.36$$

$$= \prod_{i=1}^n \left\{ f(\theta_1, \theta_2, \dots, \theta_k | x_i) \right\}^{\delta_i} \left\{ 1 - F(\theta_1, \theta_2, \dots, \theta_k | x_i) \right\}^{1-\delta_i} \quad 4.37$$

Para a distribuição de Weibull, a função de verosimilhança para dados censurados à direita tipo 1 é dada por,

$$L(\eta, \beta) = \prod_{i=1}^n \left\{ \frac{\beta}{\eta} \left(\frac{t_i}{\eta} \right)^{\beta-1} \exp \left(- \left(\frac{t_i}{\eta} \right)^\beta \right) \right\}^{\delta_i} \left\{ \exp \left(- \left(\frac{t_i}{\eta} \right)^\beta \right) \right\}^{1-\delta_i} \quad 4.38$$

Ao aplicar o logaritmo à equação anterior, vem,

$$\ln L(\eta, \beta) = l(\eta, \beta) = \sum_{i=1}^n (n\delta_i \ln \beta - n\beta\delta_i \ln \eta) + (\beta - 1) \sum_{i=1}^n (\delta_i \ln t_i) - \sum_{i=1}^n \left(\frac{t_i}{\eta} \right)^\beta \quad 4.39$$

De modo a obter os pontos máximos da função de verosimilhança é necessário resolver as derivadas parciais da equação anterior e igualar a zero, conforme indicado nas seguintes equações,

$$\frac{\partial l(\eta, \beta)}{\partial \eta} = \sum_{i=1}^n \left(-\frac{n\beta\delta_i}{\eta} \right) + \frac{\beta}{\eta} \sum_{i=1}^n \left(\frac{t_i}{\eta} \right)^\beta = 0 \quad 4.40$$

$$\frac{\partial l(\eta, \beta)}{\partial \beta} = \sum_{i=1}^n \left(\frac{n\delta_i}{\beta} - n\delta_i \ln \eta \right) + \sum_{i=1}^n (\delta_i \ln t_i) - \sum_{i=1}^n \left[\left(\frac{t_i}{\eta} \right)^\beta \ln \left(\frac{t_i}{\eta} \right) \right] = 0 \quad 4.41$$

Assim as soluções encontradas para obter a estimação dos parâmetros η e β da distribuição são dadas pelas seguintes equações,

$$\hat{\eta} = \left[\sum_{i=1}^n \left(\frac{1}{n\delta_i} t_i^{\hat{\beta}} \right) \right]^{\frac{1}{\hat{\beta}}} \quad 4.42$$

$$\sum_{i=1}^n \left(\frac{\delta_i}{\beta} + \ln \eta (1 - \delta_i) \right) + \frac{1}{n} \sum_{i=1}^n \delta_i \ln t_i - \frac{\sum_{i=1}^n t_i^\beta \ln t_i}{\sum_{i=1}^n t_i^\beta} = 0 \quad 4.43$$

A segunda derivada deverá ser negativa para garantir que os resultados obtidos correspondam a um ponto máximo. As equações da segunda derivada são as seguintes,

$$\frac{\partial^2 l(\eta, \beta)}{\partial \eta^2} = \frac{\beta}{\eta^2} \left[\sum_{i=1}^n n\delta_i + (\beta - 1) \sum_{i=1}^n \left(\frac{t_i}{\eta} \right)^\beta \right] \quad 4.44$$

$$\frac{\partial^2 l(\eta, \beta)}{\partial \eta \partial \beta} = \frac{\partial^2 l(\eta, \beta)}{\partial \beta \partial \eta} = -\frac{1}{\eta} \left[\sum_{i=1}^n (n\delta_i) - \sum_{i=1}^n \left(\frac{t_i}{\eta} \right)^\beta - \beta \sum_{i=1}^n \left[\left(\frac{t_i}{\eta} \right)^\beta \left(\ln \left(\frac{t_i}{\eta} \right) \right) \right] \right] \quad 4.45$$

$$\frac{\partial^2 l(\eta, \beta)}{\partial \beta^2} = \sum_{i=1}^n \left(-\frac{n\delta_i}{\beta^2} \right) - \sum_{i=1}^n \left[\left(\frac{t_i}{\eta} \right)^\beta \left(\ln \left(\frac{t_i}{\eta} \right) \right)^2 \right] \quad 4.46$$

b) Método de estimação de máxima verosimilhança para dados censurados à direita tipo 2

Na censura do tipo 2 todas as unidades de um sistema são observadas até à ocorrência de um determinado número de falhas. Para este esquema de censura o número de falhas do sistema é fixo, enquanto que o tempo de observação é aleatório.

Apesar do conceito entre a censura à direita do tipo 1 e do tipo 2 ser diferente a função de verosimilhança é análoga. Assim a resolução e os respetivos resultados são iguais.

A principal diferença consiste na condição para o término do resisto do tempo de análise. Nos dados censurados à direita tipo 1 a condição é o tempo de estudo enquanto nos dados censurados à direita tipo 2 a condição é o número de falhas. Contudo ambos possuem dados incompletos à direita porque permanece em funcionamento o sistema em análise no término do estudo.

Seja $t_i = t_1, t_2, \dots, t_n$, em que r registos correspondem a tempos de falha e $(n - r)$ não correspondem a tempos de falha dentro do tempo limite t .

Assim, a caracterização dos dados observados são definidos pela variável δ_i , onde,

$$\delta_i = \begin{cases} 1, & \text{para dados não censurados} \\ 0, & \text{para dados censurados à direita} \end{cases}$$

A função de verosimilhança para dados censurados à direita tipo 2 é igual à equação 4.37 e é dada por,

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n \left\{ f(\theta_1, \theta_2, \dots, \theta_k | x_i) \right\}^{\delta_i} \left\{ 1 - F(\theta_1, \theta_2, \dots, \theta_k | x_i) \right\}^{1-\delta_i}$$

A resolução e os resultados são iguais aos apresentados para os dados censurados à direita tipo 1.

c) Método de estimação de máxima verosimilhança para dados censurados à esquerda

Os dados são considerados censurados à esquerda se o tempo de falha é anterior ao tempo do início do registo. A falha já aconteceu quando a unidade foi observada. Se a falha da unidade surge antes do início do estudo, o tempo de falha é somente conhecido após certo tempo. A falha ocorreu algum tempo antes do registo, mas não há a informação de exatamente quando.

A função de verosimilhança para dados censurados à esquerda é dada por (Meeker, 1998),

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{\delta_i=1} f(\theta_1, \theta_2, \dots, \theta_k | x_i) \prod_{\delta_i=0} F(\theta_1, \theta_2, \dots, \theta_k | x_i) \quad 4.47$$

$$= \prod_{i=1}^n \left\{ f(x_i | \theta_1, \theta_2, \dots, \theta_k) \right\}^{\delta_i} \left\{ F(x_i | \theta_1, \theta_2, \dots, \theta_k) \right\}^{1-\delta_i} \quad 4.48$$

Para a distribuição de Weibull, a função de verosimilhança para dados censurados à esquerda é dada por,

$$L(\eta, \beta) = \prod_{i=1}^n \left\{ \frac{\beta}{\eta} \left(\frac{t_i}{\eta} \right)^{\beta-1} \exp \left(- \left(\frac{t_i}{\eta} \right)^{\beta} \right) \right\}^{\delta_i} \left\{ 1 - \exp \left(- \left(\frac{t_i}{\eta} \right)^{\beta} \right) \right\}^{1-\delta_i} \quad 4.49$$

Ao aplicar o logaritmo à equação anterior, vem,

$$\begin{aligned} \ln L(\eta, \beta) = l(\eta, \beta) &= \sum_{i=1}^n (n\delta_i \ln \beta - n\beta\delta_i \ln \eta) + (\beta-1) \sum_{i=1}^n (\delta_i \ln t_i) - \delta_i \sum_{i=1}^n \left(\frac{t_i}{\eta} \right)^{\beta} + \\ &+ (1-\delta_i) \sum_{i=1}^n \ln \left\{ 1 - \exp \left[- \left(\frac{t_i}{\eta} \right)^{\beta} \right] \right\} \end{aligned} \quad 4.50$$

Ao igualar a zero as derivadas parciais da equação anterior em ordem a η e β é possível obter os pontos máximos da função de verosimilhança, conforme indicado nas seguintes equações,

$$\frac{\partial l(\eta, \beta)}{\partial \eta} = \sum_{i=1}^n \left(-\frac{n\beta\delta_i}{\eta} \right) + \frac{\beta}{\eta} \sum_{i=1}^n \left[\left(\frac{t_i}{\eta} \right)^{\beta} \frac{\delta_i \exp \left(\left(\frac{t_i}{\eta} \right)^{\beta} - 1 \right)}{\exp \left(\left(\frac{t_i}{\eta} \right)^{\beta} - 1 \right)} \right] = 0 \quad 4.51$$

$$\frac{\partial l(\eta, \beta)}{\partial \beta} = \sum_{i=1}^n \left(\frac{n\delta_i}{\beta} - n\delta_i \ln \eta \right) + \sum_{i=1}^n (\delta_i \ln t_i) + \sum_{i=1}^n \left[\left(\frac{t_i}{\eta} \right)^{\beta} \ln \left(\frac{t_i}{\eta} \right) \frac{1 - \delta_i \exp \left(\left(\frac{t_i}{\eta} \right)^{\beta} \right)}{\exp \left(\left(\frac{t_i}{\eta} \right)^{\beta} - 1 \right)} \right] = 0 \quad 4.52$$

A segunda derivada deverá ser negativa para garantir que os resultados obtidos correspondam a um ponto máximo. As equações da segunda derivada são as seguintes,

$$\frac{\partial^2 l(\eta, \beta)}{\partial \eta^2} = \frac{\beta}{\eta^2} \left[\sum_{i=1}^n (n\delta_i) + \sum_{i=1}^n \left((2\delta_i - 1)(\beta - 1) \left(\frac{t_i}{\eta} \right)^{\beta} \right) \right] \quad 4.53$$

$$\frac{\partial^2 l(\eta, \beta)}{\partial \eta \partial \beta} = \frac{\partial^2 l(\eta, \beta)}{\partial \beta \partial \eta} = \frac{1}{\eta} \left[\sum_{i=1}^n (-n\delta_i) + \sum_{i=1}^n (2\delta_i - 1) \left(\frac{t_i}{\eta} \right)^{\beta} - \beta \sum_{i=1}^n \left((2\delta_i - 1) \left(\frac{t_i}{\eta} \right)^{\beta} \ln \left(\frac{t_i}{\eta} \right) \right) \right] \quad 4.54$$

$$\frac{\partial^2 l(\eta, \beta)}{\partial \beta^2} = \sum_{i=1}^n \left(-\frac{n\delta_i}{\beta^2} \right) - \sum_{i=1}^n \left[(2\delta_i - 1) \left(\frac{t_i}{\eta} \right)^\beta \left(\ln \left(\frac{t_i}{\eta} \right) \right)^2 \right] \quad 4.55$$

b) Método de estimação de máxima verosimilhança para outros tipos de dados incompletos

De entre as outras formas de dados censurados é de destacar também os dados censurados por intervalo, devido ao facto de ocorrerem com frequência na prática quando os dados provêm do registo do tempo de falha de um equipamento ou de um componente.

Conforme referido no capítulo 2.2.2.3 a censura por intervalo acontece quando por algum motivo não foi possível observar o tempo exato da falha, mas sim a ocorrência num certo intervalo de tempo. Neste tipo de censura a falha garantidamente ocorre entre duas inspeções consecutivas, com a informação que $U_i < T_i < V_i$ (Lawless, 2003). Em que U_i e V_i correspondem à variável do tempo de inspeção anterior e posterior à ocorrência da falha respetivamente.

A função de verosimilhança para dados censurados por intervalo é dada por,

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{\delta_i=1} f(\theta_1, \theta_2, \dots, \theta_k | x_i) \prod_{\delta_i=0} [F(\theta_1, \theta_2, \dots, \theta_k | v_i) - F(\theta_1, \theta_2, \dots, \theta_k | u_i)] \quad 4.56$$

$$= \prod_{i=1}^n \left\{ f(\theta_1, \theta_2, \dots, \theta_k | x_i) \right\}^{\delta_i} \left\{ F(\theta_1, \theta_2, \dots, \theta_k | v_i) - F(\theta_1, \theta_2, \dots, \theta_k | u_i) \right\}^{1-\delta_i} \quad 4.57$$

Os dados incompletos também podem ser caracterizados pela existência de dados truncados.

Embora os dados truncados sejam semelhantes aos dados censurados, o conceito difere em aspetos importantes. Os dados podem ser considerados truncados à esquerda quando as falhas que eventualmente ocorreram antes de T_e não foram registadas e não existe qualquer informação se efetivamente ocorreu a falha.

A função de verosimilhança para dados truncados à esquerda, é dada por (Stanford, 1994), (Lawless, 2003), (Cohen, 1991)

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i \in S} \frac{f(x_i | \theta_1, \theta_2, \dots, \theta_k)}{1 - F(\tau_i^e | \theta_1, \theta_2, \dots, \theta_k)} \quad 4.58$$

Em que S corresponde aos dados truncados à esquerda e τ_i^e ao tempo truncado à esquerda.

Apesar de não ser comum no estudo da fiabilidade de equipamentos e de componentes, os dados truncados à direita surgem quando o valor da variável aleatória T , é registada apenas quando é anterior ao momento truncado à direita, T_d , isto é, os valores posteriores a T_d não são observados.

Para os dados truncados à direita a função de verosimilhança é dada por (Stanford, 1994), (Lawless, 2003), (Cohen, 1991),

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i \in S} \frac{f(x_i | \theta_1, \theta_2, \dots, \theta_k)}{F(\tau_i^d | \theta_1, \theta_2, \dots, \theta_k)} \quad 4.59$$

Em que S corresponde aos dados truncados à direita e τ_i^d ao tempo truncado à direita.

4.2.2.2.1 - Simulação de Monte Carlo

Nesta secção, a simulação pelo método de Monte Carlo realizou-se com o objetivo de avaliar o desempenho do método de máxima verosimilhança (MV) e comparar com o método dos mínimos quadrados sob diferentes combinações dos parâmetros experimentais, de acordo com as considerações referidas no anexo A.

Conforme referido na Seção 2.2.4, o enviesamento, a variância e o erro quadrático médio (EQM) dos estimadores são os critérios mais comuns para avaliar o desempenho de um método de estimação.

a) Procedimento da simulação de Monte Carlo para dados completos

O procedimento passo-a-passo da simulação para dados completos é descrito como se segue:

Passo 1: Gerar n números aleatórios t_1, t_2, \dots, t_n a partir de uma distribuição de Weibull, dado os parâmetros η_T e β_T .

Passo 2: Com os resultados gerados no passo anterior, estimar η e β , pela utilização do método de máxima verosimilhança (capítulo 4.2.2.2).

Passo 3: Repetir o passo 1 ao passo 2, M vezes (M é designado por número de iterações ou número de repetições).

Passo 4: Calcular a função pivotal (equação A.1 e A.2). Calcular o enviesamento, a variância e o erro quadrático médio do estimador.

b) Procedimento da simulação de Monte Carlo para dados censurados à direita tipo 1

O procedimento passo-a-passo da simulação para dados censurados é descrito como se segue:

Passo 1: Gerar n números aleatórios t_1, t_2, \dots, t_n a partir de uma distribuição de Weibull, dado os parâmetros η_T e β_T .

Passo 2: A partir da amostra completa t_1, t_2, \dots, t_n gerada no passo anterior selecionar aleatoriamente $n-r$ tempos censurados, denominados por $t_{c,k}$ ($k = 1, 2, \dots, n-r$). As restantes observações são tempos de falha, denominados por $t_{f,l}$ ($l = 1, 2, \dots, r$). Aos tempos $t_{c,k}$ associar o valor de $\delta_j = 0$ e aos restantes tempos, $t_{f,l}$, associar o valor de $\delta_j = 1$.

Passo 3: Misturar aleatoriamente os tempos de falha com os tempos censurados para produzir a amostra censurada.

Passo 4: Para a amostra atual, estimar η e β , pela utilização do método de máxima verosimilhança (capítulo 4.2.2.2).

Passo 5: Repetir o passo 1 ao passo 4, M vezes (M é designado por número de iterações ou número de repetições).

Passo 6: Calcular a função pivotal (equação A.1 e A.2). Calcular o enviesamento, a variância e o erro quadrático médio do estimador.

Resultado da simulação do método de máxima verosimilhança

a) Resultados da simulação com dados completos

Foi examinado pela simulação de Monte Carlo o método de máxima verosimilhança para dados completos (equação 4.31 e 4.32). A comparação com o método dos mínimos quadrados focou-se com o estimador de Bernard dado que apresentou melhores resultados, conforme indicado no capítulo 4.2.2.1.1.

Na tabela 4.8 são apresentados os parâmetros da simulação para a estimação do parâmetro de forma, β , com dados completos, nomeadamente, os valores dos parâmetros reais de η e β (denominados de η_T e β_T), o tamanho da amostra n e o número de iteração M , conforme referido no anexo A.

Tabela 4.8 - Parâmetros da simulação para a estimação do parâmetro de forma, β , com dados completos.

| Parâmetros | Valores |
|------------|----------------|
| η_T | 1 |
| β_T | 1 |
| n | 4, 7, 10 - 100 |
| M | 10000 |

Os resultados de $\hat{\beta}$ são apresentados na figura 4.11 através da função pivotal. O comportamento do enviesamento é idêntico à função pivotal.

As considerações são iguais às referidas no capítulo 4.2.2.1.1.

O valor de η_T foi sempre igual a 1 e o valor de β_T foi simulado com vários valores contudo o resultado não variou significativamente com β_T o que confirma as considerações indicadas no Anexo A, assim apenas é apresentado o resultado para $\eta_T = \beta_T = 1$ para o método de máxima verosimilhança (MV) e o método de Bernard.

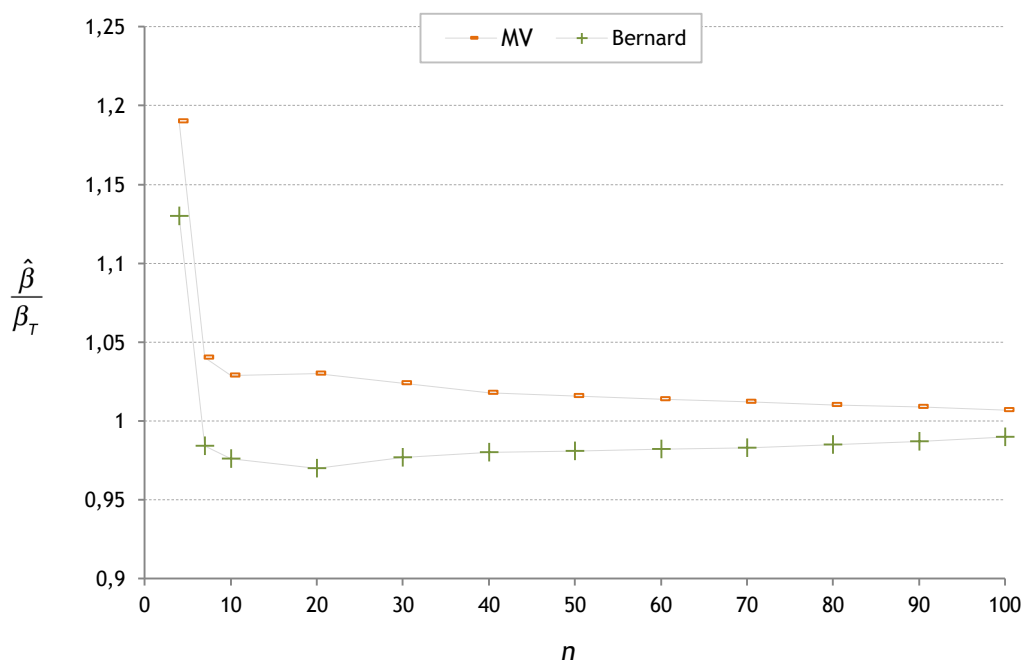


Figura 4.11 - Comparação da estimação do parâmetro de forma, β , para dados completos ($\eta_T = \beta_T = 1$).

O enviesamento relativo, $\hat{\beta} / \beta_T$, pelo método de estimação de Bernard é menor do que pelo método de máxima verosimilhança para pequenas amostras, isto é, $n \leq 20$. Contudo para amostras de maior dimensão o método de máxima verosimilhança apresenta melhores resultados.

Na tabela 4.9 são apresentados os parâmetros da simulação para a estimação do parâmetro de escala, η .

Tabela 4.9 - Parâmetros da simulação para a estimação do parâmetro de escala, η , com dados completos.

| Parâmetros | Valores |
|------------|----------------|
| η_T | 1 |
| β_T | 0.5, 1, 3 |
| n | 4, 7, 10 - 100 |
| M | 10000 |

Os resultados de $\hat{\eta}$ são apresentados da figura 4.12 à 4.14.

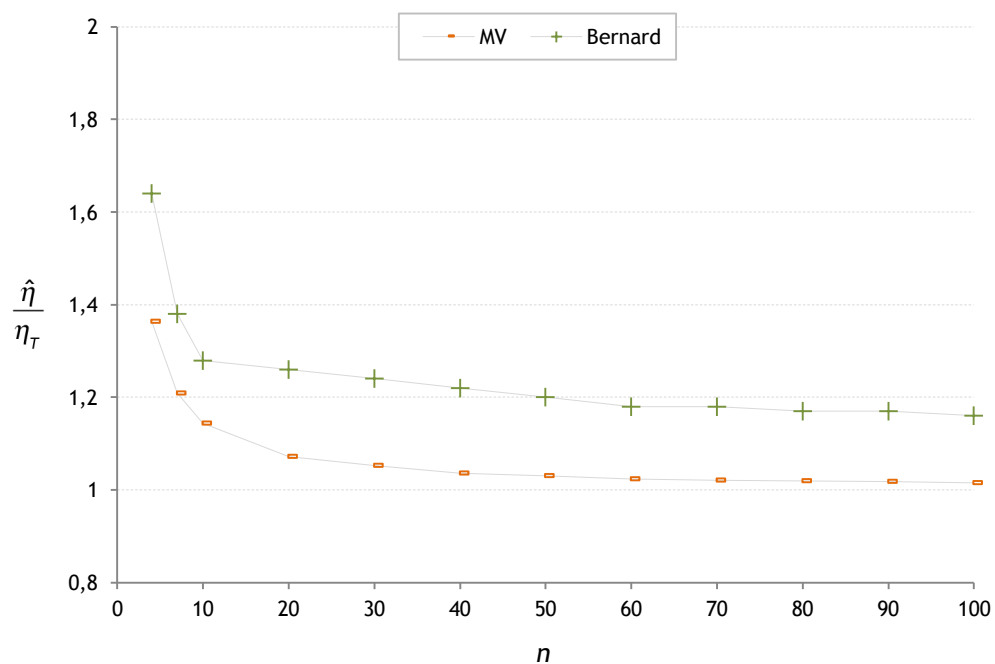


Figura 4.12 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = 1$, $\beta_T = 0.5$).

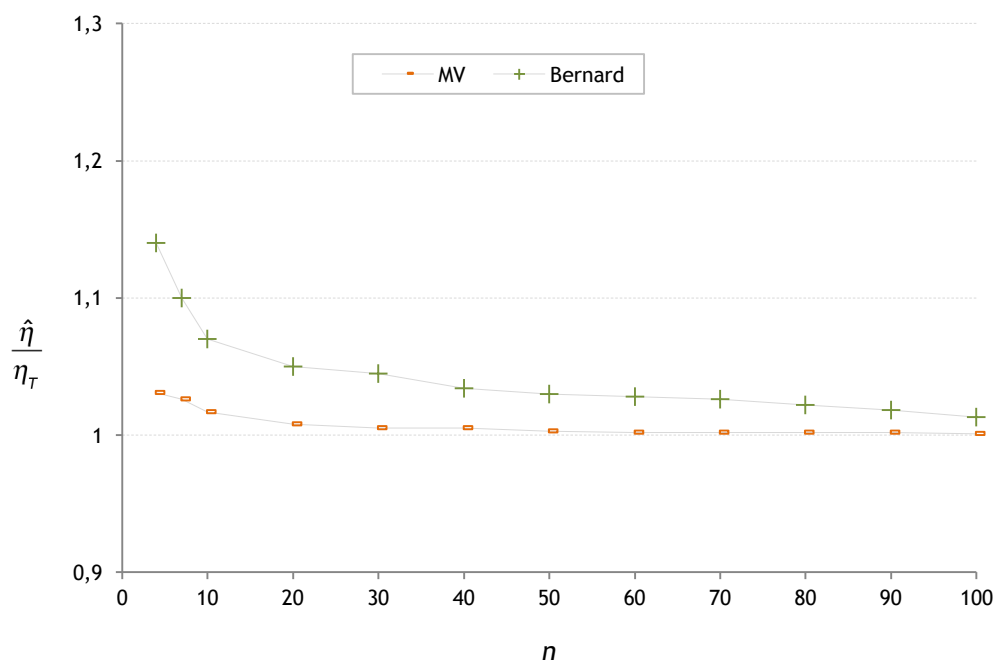


Figura 4.13 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = \beta_T = 1$).

4 - Estimação dos parâmetros da distribuição de Weibull

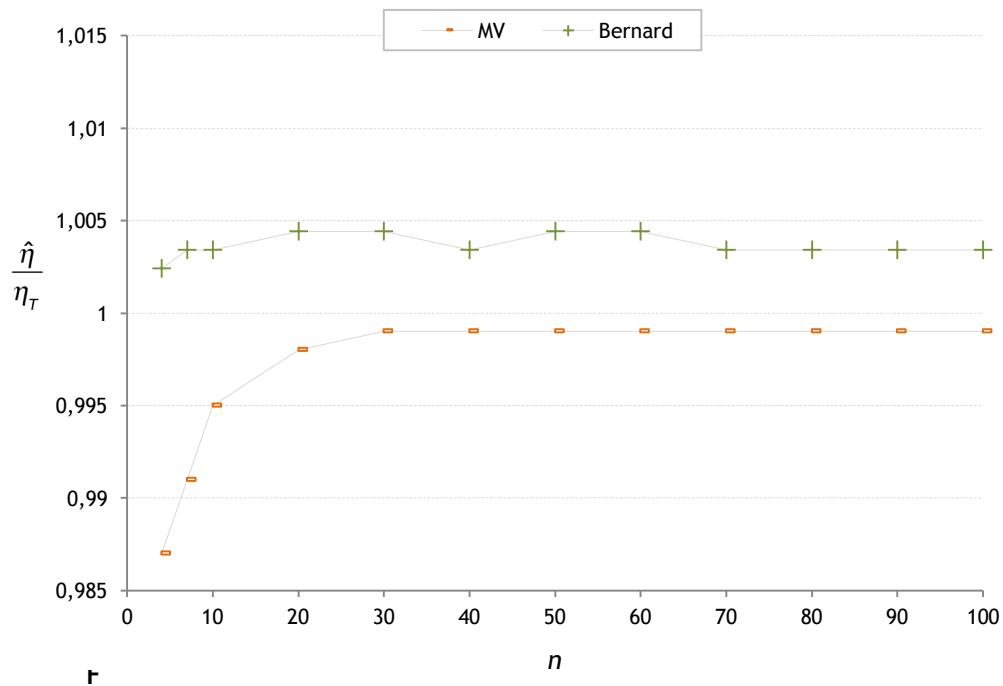


Figura 4.14 - Comparação da estimação do parâmetro de escala, η , para dados completos ($\eta_T = 1$, $\beta_T = 3$).

Comparando as três figuras (figura 4.12, 4.13 e 4.14), verifica-se que o enviesamento relativo, $\hat{\eta}/\eta_T$, diminui com o aumento de n . O método de máxima verosimilhança comparativamente ao método de estimação de Bernard apresenta melhores resultados exceto para $\beta_T = 3$ e $n < 20$.

b) Resultados da simulação com dados censurados à direita tipo 1

Foi examinado pela simulação de Monte Carlo o método de máxima verosimilhança para dados censurados à direita tipo 1 (equação 4.42 e 4.43). Apenas é apresentado o resultado da simulação para dados censurados à direita tipo 1, dado que este tipo de dado censurado é representativo da larga maioria dos casos reais e surge no caso de estudo referido no capítulo 5.

Apresenta-se o resultado em comparação com o método de estimação de Herd-Johnson dado que apresentou melhores resultados, conforme referido no capítulo 4.2.2.1.1.

Na tabela 4.10 são apresentados os parâmetros da simulação para a estimação do parâmetro de forma, β , com dados censurados, nomeadamente, os valores dos parâmetros reais de η e β (denominados de η_T e β_T), o tamanho da amostra n , percentagem de dados censurados c e o número de iteração M , conforme referido no anexo A.

Tabela 4.10 - Parâmetros da simulação para a estimação do parâmetro de forma, β , com dados censurados à direita tipo 1.

| Parâmetros | Valores |
|------------|----------------|
| η_T | 1 |
| β_T | 1 |
| n | 4, 7, 10 - 100 |
| c | 20%, 60% |
| M | 10000 |

Os resultados de $\hat{\beta}$ são apresentados na figura 4.15 e 4.16, com um nível de censura, $c=20\%$ e $c=60\%$, respectivamente, através da função pivotal. O comportamento do enviesamento é idêntico à função pivotal. As considerações são iguais às referidas no capítulo 4.2.2.1.1.

Os resultados das amostras com níveis de censura baixo (20%), figura 4.15, e das amostras com níveis de censura elevada (60%), figura 4.16, são bastantes diferentes. Nenhum dos estimadores supera os outros em todas as combinações experimentais.

O estimador Herd-Johnson apresenta um comportamento melhor quando o número de amostras é reduzido, $n < 20$.

Considerando o enviesamento e o MSE em conjunto, o método de máxima verosimilhança apresenta o melhor resultado na maioria dos casos.

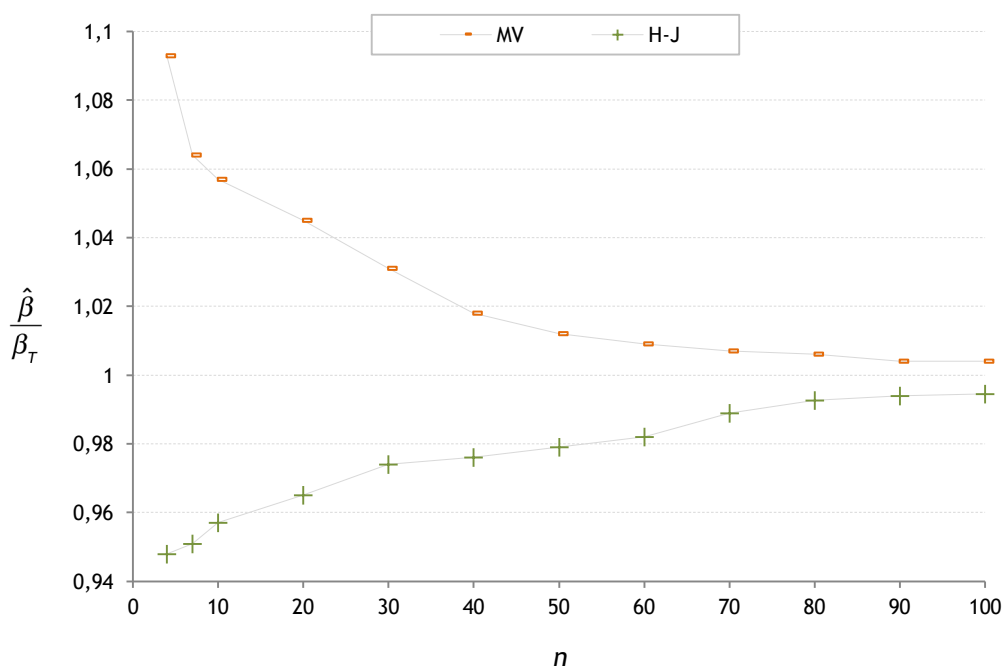


Figura 4.15 - Comparação da estimação do parâmetro de forma, β , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 20\%$.

4 - Estimação dos parâmetros da distribuição de Weibull

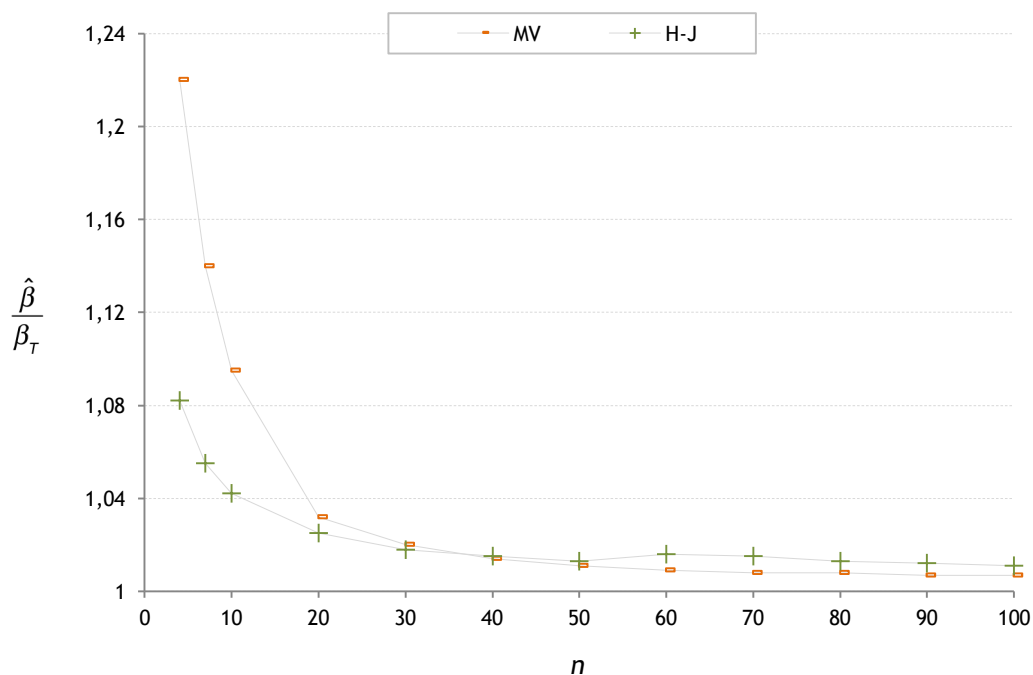


Figura 4.16 - Comparação da estimação do parâmetro de forma, β , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 60\%$.

Na tabela 4.11 são apresentados os parâmetros da simulação para a estimação do parâmetro de forma, η , com dados censurados à direita tipo 1.

Tabela 4.11 - Parâmetros da simulação para a estimação do parâmetro de forma, η , com dados censurados à direita tipo 1.

| Parâmetros | Valores |
|------------|----------------|
| η_T | 1 |
| β_T | 1, 3 |
| n | 4, 7, 10 - 100 |
| c | 20%, 60% |
| M | 10000 |

Os resultados de $\hat{\eta}$ são apresentados na figura 4.17 e 4.18, com um nível de censura, $c=20\%$ e $c=60\%$, respetivamente, para $\beta_T=1$ e na figura 4.19 e 4.20, com um nível de censura, $c=20\%$ e $c=60\%$, respetivamente, para $\beta_T=3$, através da função pivotal para o método de máxima verossimilhança (MV) e o método de Bernard.

4 - Estimação dos parâmetros da distribuição de Weibull

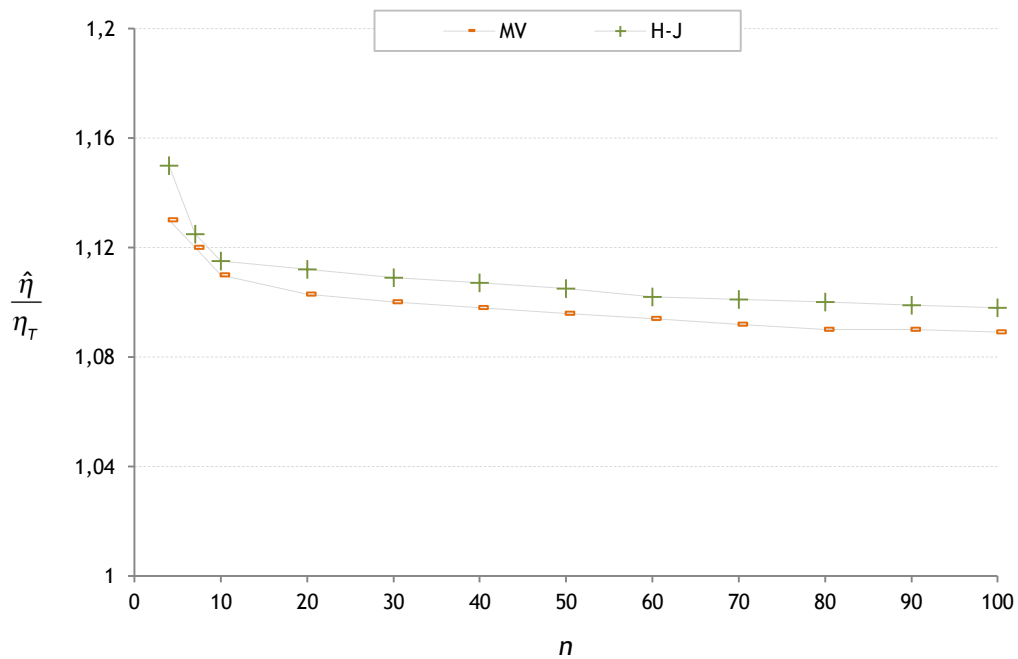


Figura 4.17 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 20\%$.

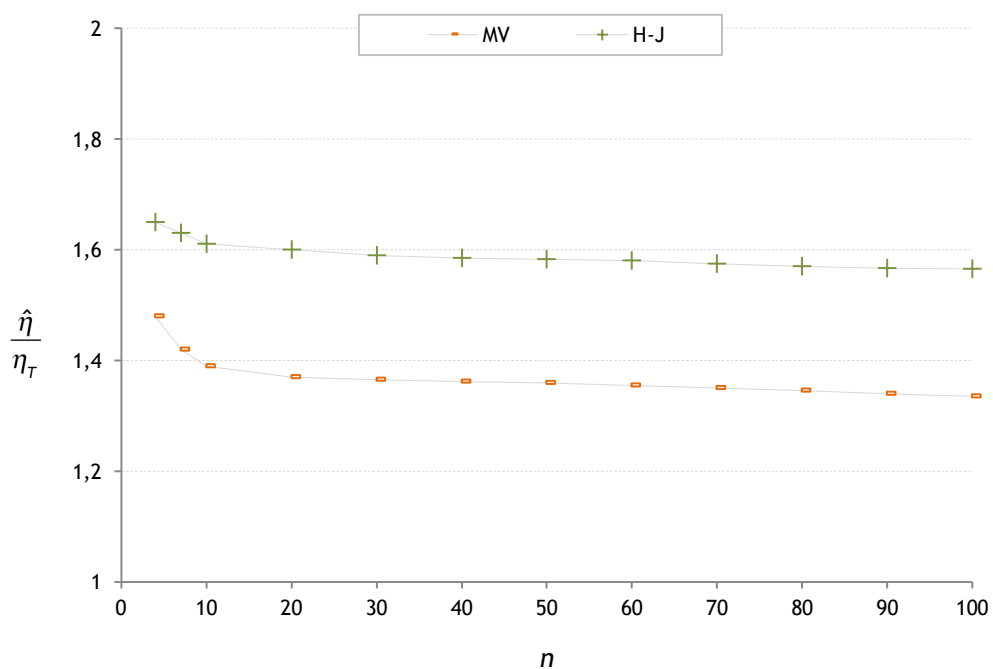


Figura 4.18 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = \beta_T = 1$), com $c = 60\%$.

4 - Estimação dos parâmetros da distribuição de Weibull

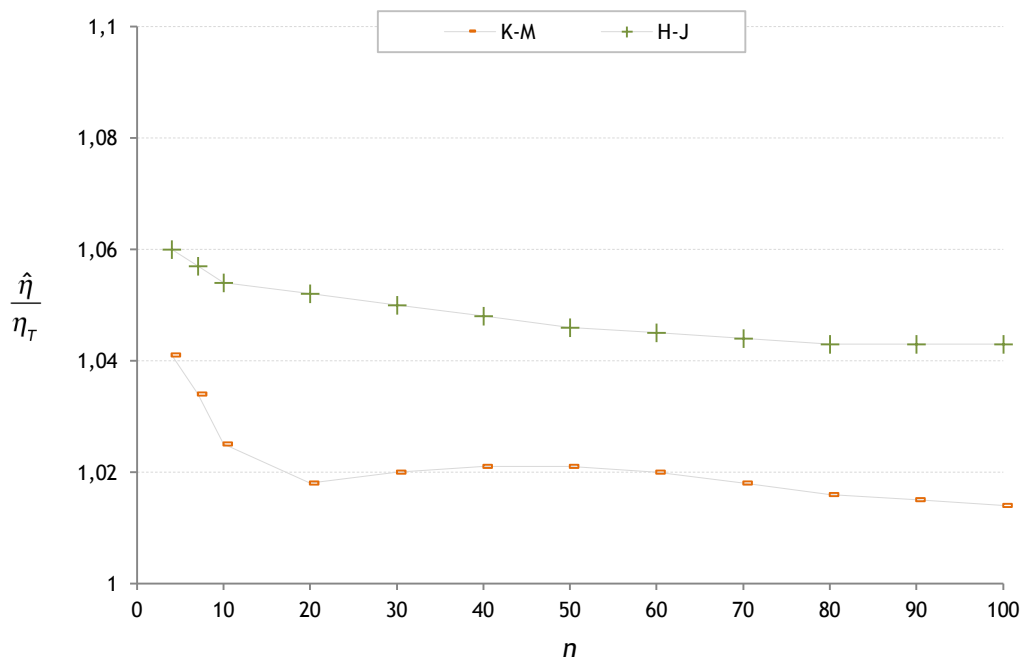


Figura 4.19 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = 1$, $\beta_T = 3$), com $c = 20\%$.

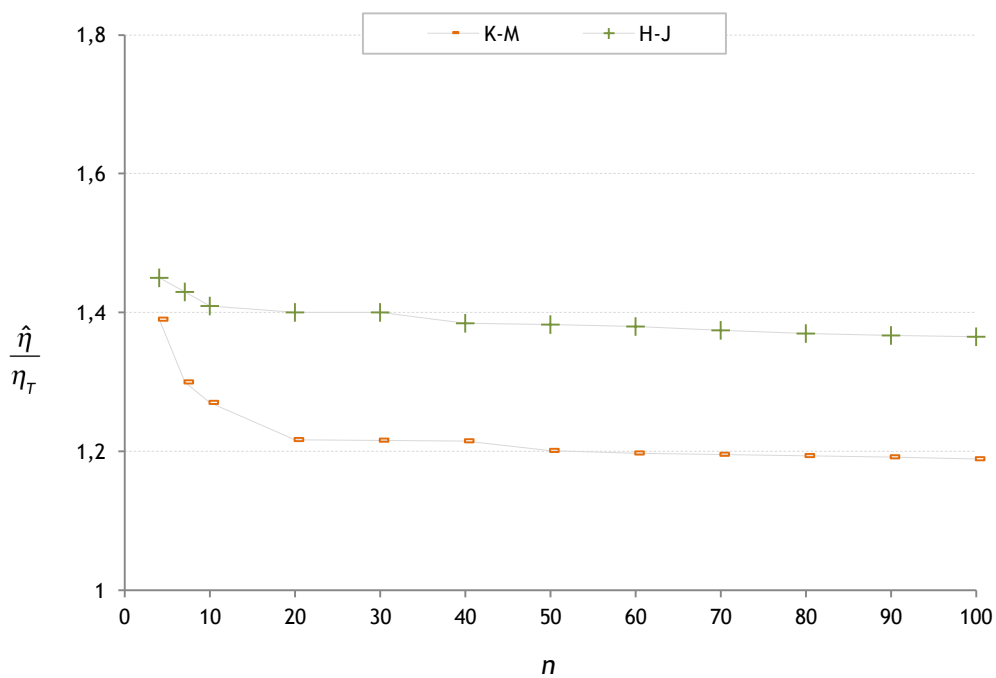


Figura 4.20 - Comparação da estimação do parâmetro de escala, η , para dados censurados ($\eta_T = 1$, $\beta_T = 3$), com $c = 60\%$.

O método de máxima verosimilhança apresenta melhores resultados na estimação de η , para dados censurados à direita tipo 1 em relação ao enviesamento relativo, $\hat{\eta} / \eta_T$, como também em relação ao erro quadrático médio (EQM) para qualquer tamanho da amostra e nível de censura.

4.2.2.2.2 - Métodos numéricos

Como as equações de máxima verosimilhança em muitas situações não apresentam solução analítica, para determinar as suas soluções, recorre-se a métodos numéricos de otimização entre os quais o método de Newton-Raphson e o algoritmo Expectation-Maximization (EM) (Kobayashi, 2012), (McLachlan, 2008).

O método de Newton-Raphson requer relativamente poucas iterações e fornece as variâncias assintóticas dos parâmetros estimados, contudo a convergência nem sempre é assegurada (Held, 2014), (McLachlan, 2008).

O algoritmo EM é de simples aplicação e a convergência monótona é assegurada, mas requer muitas iterações e pode convergir para um máximo local (Held, 2014). No entanto, este algoritmo é um dos mais eficazes na resolução das equações de máxima verosimilhança.

Dado que o algoritmo EM é utilizado no caso prático apresentado capítulo 5, é descrito com maior detalhe.

4.2.2.2.2.1 - Método de Newton-Raphson

O método de Newton-Raphson foi desenvolvido por Isaac Newton e Joseph Raphson. Em 1690, Raphson publicou uma obra, “Analysis Aequationum Universalis”, que continha um método que é agora conhecido como Método de Newton-Raphson. Contudo, em 1671, Isaac Newton já havia desenvolvido um método bastante similar, mas só foi publicada em 1736. A versão de Raphson é mais simples e por esse motivo é essa a versão que é encontrada na bibliografia atual.

O método de Newton-Raphson é um processo numérico iterativo que pode ser utilizado para estimar as raízes de uma função (McLachlan, 2008).

Um procedimento iterativo é uma técnica de aproximações sucessivas e cada aproximação é chamado uma iteração. Se as aproximações sucessivas estiverem cada vez mais próximas da solução, considera-se que as iterações convergem.

O algoritmo de Newton-Raphson, aplicado na determinação das raízes de uma função, baseia-se na expansão das funções em séries de Taylor. Seja $f(x)$ a função de interesse, a expansão em séries de Taylor da função $f(x)$ em torno de um ponto x_0 é dada por (Lee, 2003), (McLachlan, 2008):

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0) \quad 4.60$$

Em que $f'(x_0)$ é a primeira derivada de $f(x)$ para x_0 .

Para o efeito de determinar a raiz de $f(x)$, ou seja, o valor de x tal que $f(x) = 0$, dado o valor inicial x_0 , considera-se,

$$f(x_0) + (x - x_0)f'(x_0) = 0 \quad 4.61$$

A equação anterior pode ser apresentada da seguinte forma,

$$x = x_0 - \frac{f(x_0)}{f'(x_0)} \quad 4.62$$

Dando origem ao seguinte procedimento iterativo.

1. Solução inicial dos parâmetros, θ_0 ;
2. Cálculo de $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$
3. Repetir os passos até à convergência.

A convergência é alcançada quando $|x_{k+1} - x_k| < \varepsilon$, para um valor de ε pré estabelecido e provavelmente próximo de zero.

Procurando garantir a convergência do processo iterativo, esta não está sempre garantida para este método. A convergência no método de Newton-Raphson está sempre garantida para um certo intervalo $[a, b]$ que contém a raiz de $f(x)$, desde que $f(x)$ e $f'(x)$ sejam contínuas nesse intervalo. Portanto, utiliza-se uma estimativa inicial $x_0 \in [a, b]$, a convergência estará garantida. Ou seja, para o método de Newton-Raphson convergir, é preciso que a estimativa inicial esteja próxima da raiz de $f(x)$.

A escolha da solução inicial assume bastante importância na garantia e na velocidade da convergência do algoritmo.

No contexto da estimação dos parâmetros pelo método de máxima verosimilhança, pretende-se obter a maximização do logaritmo da função de verosimilhança dada pela equação 4.22. Conforme referido anteriormente a determinação dos máximos pode ser obtida pela derivada parcial do logaritmo da função de verosimilhança igual a zero.

Assim,

$$f(x) = S(\theta) = \frac{\partial l(\theta|x_i)}{\partial \theta};$$

$$f'(x) = J(\theta) = \frac{\partial^2 l(\theta|x_i)}{\partial \theta^2}, \text{ segunda derivada do logaritmo da função de verosimilhança.}$$

O método iterativo de Newton-Raphson para a estimação dos parâmetros pelo método de máxima verosimilhança, é o seguinte,

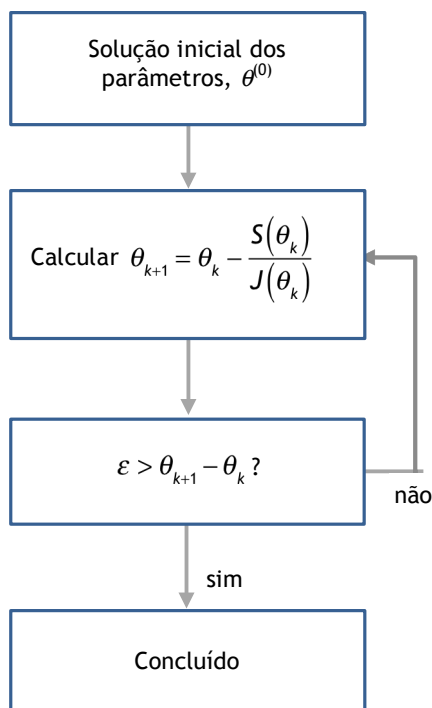


Figura 4.21 - Procedimento do método de Newton-Raphson.

4.2.2.2.2 - Algoritmo EM

O algoritmo EM (“Expectation Maximization algorithm”) é um processo iterativo que pode ser utilizado para calcular os estimadores de máxima verosimilhança em situações com dados incompletos. A designação foi dada por Dempster, Laird e Rubin (Dempster, 1977) apesar da ideia subjacente ter sido esboçada em 1972 por Orchard e Woodbury.

Anteriormente, Newcomb em 1886 e McKendrick em 1926, apresentaram trabalhos onde foi publicada metodologias semelhantes ao algoritmo EM e que também serviram de base ao trabalho publicado por Dempster, Laird e Rubin em 1977 (McLachlan, 2008).

O algoritmo consiste em dois passos realizados repetidamente até que um critério de convergência seja cumprido.

O algoritmo EM é utilizado numa vasta gama de aplicações de estatística (McLachlan, 2004), devido à sua formulação, que reduz a complexidade do problema de estimação. Como mencionado anteriormente, uma das aplicações do algoritmo é quando o estimador de máxima verosimilhança tem de ser calculado na presença de dados incompletos. O objectivo do algoritmo basicamente consiste em simplificar um problema de dados incompletos para um problema de dados completos, que muitas vezes é de mais fácil resolução, criando uma ligação entre as duas condições.

Alguns trabalhos que utilizaram o algoritmo EM na estimação dos parâmetros desconhecidos de uma distribuição têm surgido, por exemplo, Ng et al., em 2002, apresentou um estudo com dados progressivamente censurados, Balakrishnan et al., em 2011, para dados truncados à esquerda e censurados à direita para a distribuição Lognormal e em 2012, para a distribuição de Weibull, Taghipour et al., em 2013, para dados censurados por intervalo.

O algoritmo EM é um processo que converge para o estimador de máxima verosimilhança e tem como base a substituição de uma difícil maximização da verosimilhança por uma sequência de maximizações mais fáceis, cujo limite é a resposta para o problema original. A função de verosimilhança para dados completos geralmente apresenta-se de uma forma menos complexa comparativamente à função para dados incompletos (McLachlan, 2008), (Balakrishnan, 2012).

Um dos aspetos negativos do algoritmo EM é a sua convergência lenta. Contudo outros algoritmos com base no algoritmo EM têm sido propostos para aumentar a rapidez de convergência e preservando a sua simplicidade, nomeadamente o algoritmo *Incremental EM* (IEM) e o algoritmo *Sparse EM* (SPEM). A revisão destes algoritmos pode ser encontrada em Bohning (1999) e McLachlan (2000).

Seja X o conjunto de dados completos com a função densidade de probabilidade $f_c(x, \theta)$ e θ os parâmetros que caracterizam a distribuição. A função log-verosimilhança correspondente à amostra completa representa-se por,

$$\ln L_c(x, \theta) = l_c(x, \theta) \quad 4.63$$

Na presença de dados incompletos alguns eventos não são observados. Seja Y o conjunto de dados observados e Z o conjunto de dados desconhecidos, X pode ser representado em função de (y, z) , de modo que (Ng, 2002),

Cada iteração do algoritmo EM envolve dois passos, o passo E (*expectation*) e o passo M (*maximization*), definidos por (McLachlan, 2008),

Passo E: Calcular $Q(\theta | \theta^{(k)})$ onde,

$$Q(\theta, \theta^{(k)}) = E_{\theta^{(k)}} \left[l_c(x, \theta) | y, \delta, \theta^{(k-1)} \right] \quad 4.64$$

Passo M: Encontrar $\theta^{(k+1)}$ que maximiza $Q(\theta | \theta^{(k)})$ isto é,

$$\begin{aligned} \theta^{(k+1)} &= \arg \max_{\theta} Q(\theta, \theta^{(k)}) \\ Q(\theta^{(k+1)}, \theta^{(k)}) &\geq Q(\theta, \theta^{(k)}) \end{aligned} \quad 4.65$$

O procedimento é realizado até que a diferença entre a iteração k e a iteração $k+1$,

$$\varepsilon = L\left(\theta^{(k+1)}\right) - L\left(\theta^{(k)}\right), \quad 4.66$$

diminua para um valor aceitável, $\varepsilon > 0$.

O passo E do algoritmo calcula o valor esperado condicional do logaritmo da função de verosimilhança para dados completos dada a amostra observada e o passo M encontra o seu máximo.

Este algoritmo necessita de uma solução inicial para os valores dos parâmetros da distribuição, designada por $\theta^{(0)}$. A escolha desta solução inicial requer particular atenção na medida em que a velocidade de convergência do algoritmo pode tornar-se extremamente lenta devido a uma má escolha. Outro aspeto a ter em conta é que a equação de máxima verosimilhança pode ter múltiplas soluções correspondentes a máximos locais, por isso, a escolha da solução inicial torna-se importante.

Um estudo comparativo de várias estratégias na escolha dos valores iniciais foi realizado por Karlis e Xekalaki (2003). Os resultados mostram claramente a dependência da estratégia na escolha das soluções iniciais.

A figura seguinte ilustra o procedimento do algoritmo EM.

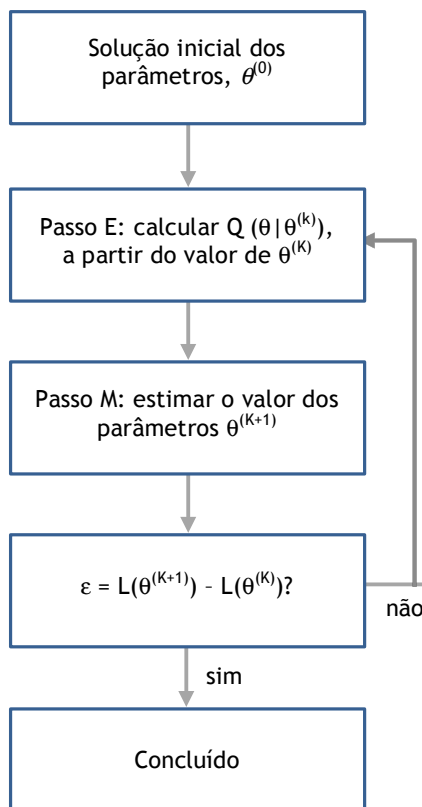


Figura 4.22 - Procedimento do algoritmo EM.

O processo iterativo é constituído por uma série de ciclos em que os dados dos parâmetros são repetidamente atualizados até que se cumpra o critério de convergência.

Quando o algoritmo EM é utilizado para estimar os parâmetros da função de máxima verosimilhança em problemas de dados incompletos, não é possível obter diretamente do algoritmo a variância e a covariância. Contudo, algumas técnicas alternativas têm sido desenvolvidas para contornar este aspeto.

A técnica utilizada neste trabalho consiste no princípio da informação desconhecida (Balakrishnan, 2012), (Ng, 2002), (Kinaci, 2014) para a obtenção da matriz da informação observada, tal que,

$$\text{Dados observados (Y)} = \text{Dados completos (X)} - \text{Dados desconhecidos (Z)}$$

Considera-se que $I_Y(\theta)$, $I_X(\theta)$ e $I_{Z|Y}(\theta)$ representa a matriz da informação observada, a matriz da informação completa e a matriz da informação desconhecida respetivamente.

A matriz da informação completa é dada por,

$$I_X(\theta) = -E \left[\frac{\partial^2}{\partial \theta^2} \log L_c(t; \theta) \right] \quad 4.67$$

A matriz da informação desconhecida para a i^{ma} observação que é censurada é dada por,

$$I_{Z|Y}^{(i)}(\theta) = -E \left[\frac{\partial^2}{\partial \theta^2} \ln f(t_i | t_i > y_i, \theta) \right] \quad 4.68$$

então,

$$I_{Z|Y}(\theta) = \sum_{i=1}^n I_{Z|Y}^{(i)}(\theta) \quad 4.69$$

Assim, pelo princípio da informação desconhecida, a matriz da informação observada pode ser obtida por,

$$I_Y(\theta) = I_X(\theta) - I_{Z|Y}(\theta) \quad 4.70$$

A matriz da variância/ covariância da função de máxima verosimilhança para θ , pode ser obtida pela inversão da matriz da informação observada.

$$\begin{bmatrix} \text{Var}(\hat{\eta}) & \text{Cov}(\hat{\eta}, \hat{\beta}) \\ \text{Cov}(\hat{\eta}, \hat{\beta}) & \text{Var}(\hat{\beta}) \end{bmatrix} = \begin{bmatrix} I_{Y,\eta\eta} & I_{Y,\eta\beta} \\ I_{Y,\beta\eta} & I_{Y,\beta\beta} \end{bmatrix}^{-1} \quad 4.71$$

Onde $I_{\gamma, \eta\eta}$, $I_{\gamma, \beta\beta}$ e $I_{\gamma, \eta\beta} = I_{\gamma, \beta\eta}$ são os elementos da matriz da informação observada.

O algoritmo EM tem várias propriedades que o destacam relativamente a outros algoritmos iterativos nomeadamente o método de Newton-Raphson:

- O algoritmo EM converge sob condições bastante amplas, ou seja, a partir de um dado arbitrário, $\theta^{(0)}$, o algoritmo geralmente encontra um máximo local, com exceção de uma má escolha da solução inicial $\theta^{(0)}$ ou na formulação errada da função de verosimilhança.
- O trabalho analítico necessário é mais simples do que com outros métodos, dado que só é necessário maximizar o valor esperado condicional da log-verosimilhança para dados completos.
- O algoritmo EM é relativamente fácil de programar e de ser implementado.
- Durante as iterações é possível controlar a convergência e os erros de programação

Contudo também apresenta algumas desvantagens:

- O algoritmo EM pode convergir lentamente, mesmo em alguns problemas aparentemente simples e nos problemas em que há muita informação incompleta.
- O algoritmo EM não tem um processo integrado para produzir uma estimativa da matriz de covariância dos parâmetros estimados. Contudo esta desvantagem pode ser contornada pela utilização de metodologia adequada.
- O algoritmo EM, como o método de Newton-Raphson, não garante a convergência para o máximo global quando existem vários máximos locais. A estimativa obtida depende da solução inicial.

4.2.2.2.2.1 - Algoritmo EM com dados censurados à direita

Um problema com tempos de falhas censurados à direita pode ser visto como um problema de dados incompletos, então o algoritmo EM é aplicável para obter a estimação dos parâmetros da função de máxima verosimilhança.

Muitos dos casos reais de dados incompletos correspondem a dados censurados à direita. Este tipo de dado censurado será aplicado no capítulo 5.

O autor desta tese utilizou o procedimento base do algoritmo EM e aplicou-o no caso específico de dados censurados à direita provenientes de um registo histórico, de forma diferente do habitual, tornando este procedimento original e distinto dos trabalhos publicados por outros autores (Balakrishnan, 2007), (Balakrishnan, 2012). Nos referidos trabalhos o contexto da aplicação é referente a uma análise de sobrevivência, onde é registado o número de componentes que falharam até ao instante Cd . No trabalho proposto a contexto é diferente. O componente avaria e é substituído por outro equipamento idêntico e é feito o registo de falhas até à conclusão do período de análise, Cd . Assim, os dados

censurados à direita verificam-se no último componente em funcionamento quando termina o tempo de registo.

Como o sistema em análise é diferente dos trabalhos anteriormente publicados optou-se por uma abordagem diferenciada conforme descrito neste capítulo.

O vetor dos dados, Z_i , pode ser considerado como os dados não observados. Os dados observados consistem em (Y_i, δ_i) , em $Y_i = \min(T_i, C_i)$, onde C_i corresponde ao tempo de observação, $\delta_i = 1$ ($T_i \leq C_i$), dado não censurado e $\delta_i = 0$ ($T_i > Y_i$), dado censurado.

O passo E do algoritmo requer o cálculo do valor esperado condicional da log-verosimilhança para dados completos dada a amostra observada. Neste caso a função log-verosimilhança para dados completos é dada pela equação 4.28, para n dados recolhidos,

$$\ln L(\eta, \beta) = l(\eta, \beta) = n \ln \beta - n \beta \ln \eta + (\beta - 1) \sum_{i=1}^n (\ln t_i) - \sum_{i=1}^n \left(\frac{t_i}{\eta} \right)^\beta$$

Considera-se que $\theta = (\eta, \beta)$ os parâmetros da distribuição, $\delta_i = (\delta_1, \delta_2, \dots, \delta_n)$ o vetor indicador de censura e $y_i = (y_1, y_2, \dots, y_n)$ o vetor dos dados observados.

Da equação 4.64, vem,

$$\begin{aligned} Q(\theta, \theta^{(k)}) &= E_{\theta^{(k)}} \left[l_c(x, \theta) \middle| y, \delta, \theta^{(k-1)} \right] \\ &= n \ln \beta - n \beta \ln \eta + (\beta - 1) \sum_{i=1}^n A_i^{(s)} - \frac{1}{\eta^\beta} \sum_{i=1}^n B_i^{(s)} \end{aligned} \quad 4.72$$

Em que,

$$A_i^{(s)} = E_{\theta^{(k)}} \left[\ln t_i \middle| y, \delta \right] = \delta_i \ln y_i + (1 - \delta_i) E_{\theta^{(k)}} \left[\ln t_i \middle| t_i > y_i \right] \quad 4.73$$

$$B_i^{(s)} = E_{\theta^{(k)}} \left[t_i^\beta \middle| y, \delta \right] = \delta_i y_i^\beta + (1 - \delta_i) E_{\theta^{(k)}} \left[t_i^\beta \middle| t_i > y_i \right] \quad 4.74$$

A expressão geral do valor esperado condicional é dada por,

$$E \left[y \middle| Y = y' \right] = \int_{y'}^{\infty} y \cdot f(y | y') dy \quad 4.75$$

Para obter o valor esperado condicional é necessário primeiro determinar a expressão da função de densidade de probabilidade condicional correspondente, $f_{(y|y')}(y|y')$. Para o caso em análise é dada por (Ng, 2002), (Nelson, 1982),

$$f(t_i | y_i) = \frac{f(t_i)}{1 - F(y_i)} \quad 4.76$$

$$= \frac{\beta}{\eta} \left(\frac{t_i}{\eta} \right)^{\beta-1} \exp \left[\left(\frac{y_i}{\eta} \right)^{\beta} - \left(\frac{t_i}{\eta} \right)^{\beta} \right], \quad t_i > y_i \quad 4.77$$

Assim,

$$\begin{aligned} E_{\theta^{(k)}} \left[\ln t_i | t > y_i \right] &= \int_{y_i}^{\infty} \left[\ln t_i \frac{\beta^{(k)}}{\eta^{(k)}} \left(\frac{t_i}{\eta^{(k)}} \right)^{\beta^{(k)}-1} \exp \left[\left(\frac{y_i}{\eta^{(k)}} \right)^{\beta^{(k)}} - \left(\frac{t_i}{\eta^{(k)}} \right)^{\beta^{(k)}} \right] \right] dt_i \\ &= \ln y_i + \frac{1}{\beta^{(k)}} \exp \left(\frac{y_i}{\eta^{(k)}} \right)^{\beta^{(k)}} \Gamma \left[0, \left(\frac{y_i}{\eta^{(k)}} \right)^{\beta^{(k)}} \right] \end{aligned} \quad 4.78$$

e,

$$\begin{aligned} E_{\theta^{(k)}} \left[t_i^{\beta} | t > y_i \right] &= \int_{y_i}^{\infty} \left[t_i^{\beta^{(k)}} \frac{\beta^{(k)}}{\eta^{(k)}} \left(\frac{t_i}{\eta^{(k)}} \right)^{\beta^{(k)}-1} \exp \left[\left(\frac{y_i}{\eta^{(k)}} \right)^{\beta^{(k)}} - \left(\frac{t_i}{\eta^{(k)}} \right)^{\beta^{(k)}} \right] \right] dt_i \\ &= y_i^{\beta^{(k)}} + \left(\eta^{(k)} \right)^{\beta^{(k)}} \end{aligned} \quad 4.79$$

onde,

$$\Gamma(p, x) = \int_x^{\infty} u^{p-1} e^{-u} du, \quad \text{é a função gama incompleta.}$$

Substituindo as equações 4.78 e 4.79 nas equações 4.73 e 4.74 respetivamente, vem,

$$A_i^{(s)} = \delta_i \ln y_i + (1 - \delta_i) \left\{ \ln y_i + \frac{1}{\beta^{(k)}} \exp \left(\frac{y_i}{\eta^{(k)}} \right)^{\beta^{(k)}} \Gamma \left[0, \left(\frac{y_i}{\eta^{(k)}} \right)^{\beta^{(k)}} \right] \right\} \quad 4.80$$

$$B_i^{(s)} = \delta_i y_i^{\beta} + (1 - \delta_i) \left[y_i^{\beta^{(k)}} + \left(\eta^{(k)} \right)^{\beta^{(k)}} \right] \quad 4.81$$

Assim, a expressão referente ao passo E do algoritmo EM, $Q(\theta, \theta^{(k)})$, com dados censurados à direita é dada pela seguinte equação,

$$Q(\theta, \theta^{(k)}) = n \ln \beta - n \beta \ln \eta + (\beta - 1) \sum_{i=1}^n \left[\delta_i \ln y_i + (1 - \delta_i) \left\{ \ln y_i + \frac{1}{\beta^{(k)}} \exp \left(\frac{y_i}{\eta^{(k)}} \right)^{\beta^{(k)}} \Gamma \left[0, \left(\frac{y_i}{\eta^{(k)}} \right)^{\beta^{(k)}} \right] \right\} \right] -$$

$$-\frac{1}{\eta^\beta} \sum_{i=1}^n \left\{ \delta_i y_i^\beta + (1-\delta_i) \left[y_i^{\beta^{(k)}} + \left(\eta^{(k)} \right)^{\beta^{(k)}} \right] \right\} \quad 4.82$$

Conforme referido anteriormente, no passo M pretende-se encontrar a solução $\theta^{(k+1)}$ que maximiza $Q(\theta, \theta^{(k)})$. De modo a obter os pontos máximos é necessário resolver as derivadas parciais da equação anterior e igualar a zero, conforme indicado nas seguintes equações,

$$\frac{\partial Q(\theta, \theta^{(k)})}{\partial \eta} = -\frac{n\beta}{\eta} + \frac{\beta}{\eta^{\beta+1}} \sum_{i=1}^n \left\{ \delta_i y_i^\beta + (1-\delta_i) \left[y_i^{\beta^{(k)}} + \left(\eta^{(k)} \right)^{\beta^{(k)}} \right] \right\} \quad 4.83$$

$$\begin{aligned} \frac{\partial Q(\theta, \theta^{(k)})}{\partial \beta} = & \frac{n}{\beta} - n \ln \eta + \sum_{i=1}^n \left[\delta_i \ln y_i + (1-\delta_i) \left\{ \ln y_i + \frac{1}{\beta^{(k)}} \exp \left(\frac{y_i}{\eta^{(k)}} \right)^{\beta^{(k)}} \Gamma \left[0, \left(\frac{y_i}{\eta^{(k)}} \right)^{\beta^{(k)}} \right] \right\} \right] + \\ & + \sum_{i=1}^n \left[\frac{\ln \eta}{\eta^\beta} \left\{ \delta_i y_i^\beta + (1-\delta_i) \left[y_i^{\beta^{(k)}} + \left(\eta^{(k)} \right)^{\beta^{(k)}} \right] \right\} - \frac{1}{\eta^\beta} \delta_i y_i^\beta \ln y_i \right] \end{aligned} \quad 4.84$$

Assim a solução encontrada para obter a estimação dos parâmetros η da distribuição é dada pela seguinte equação,

$$\eta = \sum_{i=1}^n \left\{ \delta_i y_i^\beta + (1-\delta_i) \left(y_i^{\beta^{(k)}} + \eta^{(k)} \exp \left(\frac{y_i}{\eta^{(k)}} \right)^{\beta^{(k)}} \right) \right\}^{\frac{1}{\beta}} \cdot n^{-\frac{1}{\beta}} \quad 4.85$$

Com a equação 4.85 encontrada para o parâmetro η , por substituição é possível obter a solução referente ao parâmetro β da distribuição.

A segunda derivada deverá ser negativa para garantir que os resultados obtidos correspondam a um ponto máximo. As equações da segunda derivada são as seguintes,

$$\frac{\partial^2 Q(\theta, \theta^{(k)})}{\partial \eta^2} = \frac{n\beta}{\eta^2} - \frac{\beta(\beta+1)}{\eta^{\beta+2}} \sum_{i=1}^n \left\{ \delta_i y_i^\beta + (1-\delta_i) \left[y_i^{\beta^{(k)}} + \left(\eta^{(k)} \right)^{\beta^{(k)}} \right] \right\} \quad 4.86$$

$$\begin{aligned} \frac{\partial^2 Q(\theta, \theta^{(k)})}{\partial \beta^2} = & -\frac{n}{\beta^2} - \frac{(\ln \eta)^2}{\eta^\beta} \sum_{i=1}^n \left\{ \delta_i y_i^\beta + (1-\delta_i) \left[y_i^{\beta^{(k)}} + \left(\eta^{(k)} \right)^{\beta^{(k)}} \right] \right\} + \frac{2 \ln \eta}{\eta^\beta} \sum_{i=1}^n (\delta_i y_i^\beta \ln y_i) - \\ & - \frac{1}{\eta^\beta} \sum_{i=1}^n \left[\delta_i y_i^\beta (\ln y_i)^2 \right] \end{aligned} \quad 4.87$$

$$\begin{aligned} \frac{\partial^2 Q(\theta, \theta^{(k)})}{\partial \eta \partial \beta} &= \frac{\partial^2 Q(\theta, \theta^{(k)})}{\partial \beta \partial \eta} = -\frac{n}{\eta} + \frac{1-\beta \ln \eta}{\eta^{\beta+1}} \sum_{i=1}^n \left\{ \delta_i y_i^\beta + (1-\delta_i) \left[y_i^{\beta^{(k)}} + \left(\eta^{(k)} \right)^{\beta^{(k)}} \right] \right\} + \\ &+ \frac{\beta}{\eta^{\beta+1}} \sum_{i=1}^n (\delta_i y_i^\beta \ln y_i) \end{aligned} \quad 4.88$$

Conforme referido anteriormente o algoritmo EM não permite determinar a variância e a covariância relativamente à estimação dos parâmetros da função de máxima verosimilhança. Contudo a técnica apresentada permite suprimir esse problema.

Os elementos da matriz da informação completa, $I_X(\eta, \beta)$ são dados conforme referido anteriormente, pelo valor esperado, (Rinne, 2009).

$$-E \left[\frac{\partial^2}{\partial \eta^2} \log L_c(t; \theta) \right] = \frac{n\beta^2}{\eta^2} \quad 4.89$$

$$-E \left[\frac{\partial^2}{\partial \beta^2} \log L_c(t; \theta) \right] = -\frac{n}{\eta} \Gamma'(2) \approx -0,422784 \frac{n}{\eta} \quad 4.90$$

$$-E \left[\frac{\partial^2}{\partial \eta \partial \beta} \log L_c(t; \theta) \right] = \frac{n}{\beta^2} [1 + \Gamma''(2)] \approx 1.823680 \frac{n}{\beta^2} \quad 4.91$$

onde,

$$\Gamma'(x) = \Psi(x) \cdot \Gamma(x)$$

$$\Gamma''(x) = [\Psi^2(x) + \Psi'(x)] \cdot \Gamma(x)$$

$\Psi(x)$ - função digama

$\Psi'(x)$ - função trigama

Para obter os elementos da matriz da informação desconhecida, $I_{Z|X}(\eta, \beta)$ é necessário primeiro considerar o logaritmo da função de densidade de probabilidade condicional referente a dados truncados. A partir da equação 4.77, vem,

$$\ln f_{t_i|y_i} \left(t_i | t_i > y_i, \eta, \beta \right) = \ln \frac{\beta}{\eta} + \ln \left(\frac{t_i}{\eta} \right)^{\beta-1} + \left(\frac{y_i}{\eta} \right)^\beta - \left(\frac{t_i}{\eta} \right)^\beta \quad 4.92$$

Da equação anterior, obtêm-se as seguintes expressões,

$$\frac{\partial^2}{\partial \eta^2} (\ln f(t_i | t_i > y_i)) = \frac{1}{\eta^2} + \frac{\beta-1}{\eta^2} + \frac{\beta}{\eta^2} \left[\left(\frac{y_i}{\eta} \right)^\beta - \left(\frac{t_i}{\eta} \right)^\beta \right] + \frac{\beta^2}{\eta^2} \left[\left(\frac{y_i}{\eta} \right)^\beta - \left(\frac{t_i}{\eta} \right)^\beta \right] \quad 4.93$$

$$\frac{\partial^2}{\partial \beta^2} (\ln f(t_i | t_i > y_i)) = -\frac{1}{\beta^2} + \left(\frac{y_i}{\eta} \right)^\beta \left(\ln \left(\frac{y_i}{\eta} \right) \right)^2 - \left(\frac{t_i}{\eta} \right)^\beta \left(\ln \left(\frac{t_i}{\eta} \right) \right)^2 \quad 4.94$$

$$\begin{aligned} \frac{\partial^2}{\partial \eta \partial \beta} (\ln f(t_i | t_i > y_i)) &= \frac{\partial^2}{\partial \beta \partial \eta} (\ln f(t_i | t_i > y_i)) = \frac{1}{\eta} \left[-1 - \left(\frac{y_i}{\eta} \right)^\beta + \left(\frac{t_i}{\eta} \right)^\beta \right] + \\ &+ \frac{\beta}{\eta} \left[-\left(\frac{y_i}{\eta} \right)^\beta \ln \left(\frac{y_i}{\eta} \right) + \left(\frac{t_i}{\eta} \right)^\beta \ln \left(\frac{t_i}{\eta} \right) \right] \end{aligned} \quad 4.95$$

Para determinar a matriz da informação desconhecida, é necessário obter o valor esperado das equações anteriores. Verifica-se que as funções de interesse são dadas pelas equações 4.78 e 4.79, ou seja,

$$E \left[\ln t_i | t_i > y_i \right] = \ln y_i + \frac{1}{\beta} \exp \left(\frac{y_i}{\eta} \right)^\beta \Gamma \left[0, \left(\frac{y_i}{\eta} \right)^\beta \right] \quad 4.96$$

$$E \left[t_i^\beta | t_i > y_i \right] = y_i^\beta + \eta^\beta \quad 4.97$$

A partir das equações 4.93 a 4.95 e das equações 4.96 e 4.97, a matriz da informação desconhecida $I_{Z|X}(\eta, \beta)$, pode ser obtida pela equação 4.69.

Finalmente, pelo princípio da informação desconhecida, a matriz da informação observada pode ser obtida pela equação 4.70,

$$I_Y(\theta) = I_X(\theta) - I_{Z|Y}(\theta)$$

A matriz da variância/ covariância da função de máxima verosimilhança para η e β , pode ser obtida pela equação 4.71.

Depois de determinar a estimação dos parâmetros da função de máxima verosimilhança e a variância, é possível obter o intervalo de confiança para η e β .

4.3 - Intervalo de confiança

Uma estimativa pontual de um parâmetro de uma distribuição tal como apresentado anteriormente, é um número que se encontra na vizinhança do verdadeiro e desconhecido valor do parâmetro. A questão do erro presente na estimação pontual dos parâmetros devido

à variabilidade inerente à aleatoriedade das amostras que lhe deram origem remete à construção dos intervalos de confiança.

Pode-se interpretar o intervalo de confiança como um intervalo que contém os valores prováveis que o parâmetro pode assumir. Assim, a amplitude do intervalo está associada à incerteza que se tem a respeito do parâmetro. Geralmente tem-se o valor num intervalo com coeficiente de confiança $(1-\alpha)$, isso significa que o intervalo deve oferecer 100 $(1-\alpha)$ % de confiança. Por exemplo, se $\alpha = 0,05$, então o intervalo resultante é muitas vezes chamado de intervalo de confiança de 95%. Neste contexto, o coeficiente de confiança (expresso em percentagem) é muitas vezes referido como o nível de confiança.

Os intervalos de confiança podem ser bilaterais ou unilaterais (O'Connor, 2012).

- a) Bilateral: quando se aplica limites de confiança dos dois lados (ou intervalos), resulta num intervalo fechado onde uma determinada percentagem da população é provável que se encontre. Ou seja, determinam-se os valores ou limites, entre as quais se encontra uma determinada percentagem da população. Por exemplo, quando se aplica 90% de confiança bilateral, significa que 90% da população está entre X e Y com 5% menor que X e 5% maior que Y, conforme a figura seguinte.

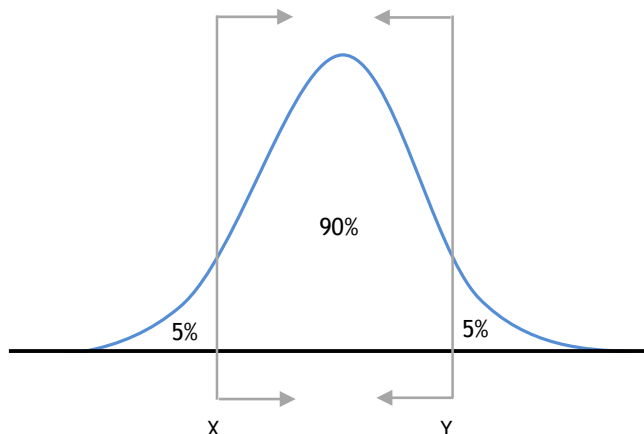


Figura 4.23 - Intervalo de confiança bilateral.

- b) Unilateral: um limite unilateral define o ponto em que uma certa percentagem da população é maior ou menor do que o ponto definido. Isto significa que existem dois tipos de limites de um lado superior e inferior. Um limite superior unilateral define um ponto em que uma certa percentagem da população é inferior a X. Por outro lado, um limite inferior unilateral define um ponto que uma determinada percentagem da população é maior do que X. Por exemplo, 95% de confiança unilateral significa que 95% da amostra é maior que X, sendo X o limite inferior, ou 95% da amostra é menor que X, sendo X o limite superior, conforme a figura seguinte.

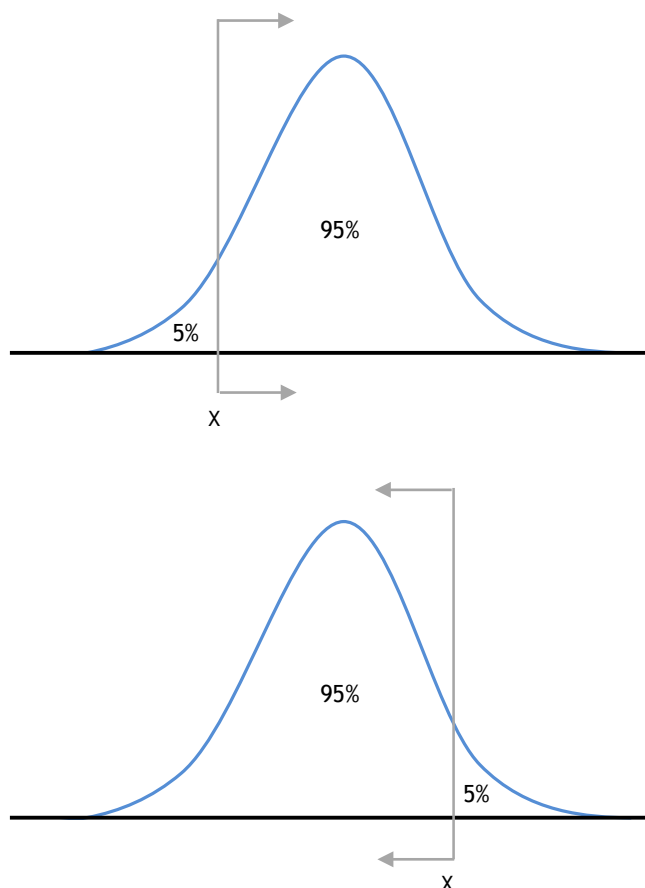


Figura 4.24 - Intervalo de confiança unilateral inferior e superior.

Os limites bilaterais significam que, dado um nível de confiança, o valor real estará entre esses limites, enquanto que no limite unilateral inferior $X\%$ de certeza que o valor é maior do que a linha limite e no caso do limite unilateral superior $X\%$ de certeza que o valor é menor do que a linha limite.

Existem muito métodos para estimar o intervalo de confiança (Abernethy, 2006), (Lawless, 2003), (Meeker, 1998), (O'Connor, 2012). Neste trabalho são apresentados três métodos, nomeadamente, método da matriz de Fisher, método da razão de verossimilhança e método de Bootstrap.

Utiliza-se em todos os intervalos de confiança mencionados a notação $[L_l(\theta) ; L_s(\theta)]$ em que $L_l(\theta)$ é o limite inferior do intervalo e $L_s(\theta)$ é o limite superior do intervalo.

4.3.1 - Método da matriz de Fisher

A estimação do intervalo de confiança pelo método da matriz de Fisher é amplamente utilizada em muitas aplicações estatísticas. O intervalo de confiança é calculado através da informação obtida pela matriz de Fisher.

Conforme referido anteriormente, o inverso da matriz de Fisher produz a matriz de variância-covariância, que fornece a informação do parâmetro da variância, $Var(\hat{\theta})$ (Panahi, 2011).

Considera-se um coeficiente de confiança 100 (1- α) % para a estimativa do parâmetro e encontrado z_c na tabela da distribuição normal, assim, o intervalo de confiança pelo método da matriz de Fisher é calculado pelas seguintes equações,

$$\begin{aligned} L_l(\theta) &= \hat{\theta} - z_c \cdot \sqrt{Var(\hat{\theta})} \\ L_s(\theta) &= \hat{\theta} + z_c \cdot \sqrt{Var(\hat{\theta})} \end{aligned} \quad 4.98$$

É mais adequado aplicar este método em amostras que se ajustam à distribuição normal ou para amostras de grande dimensão. Dado que, o teorema do limite central afirma que quando o tamanho da amostra, $n \rightarrow \infty$ a amostra tende a ajustar-se a uma distribuição normal.

Para amostras de pequena dimensão é mais recomendável que se utilize outro método (Abernethy, 2006).

4.3.2 - Método da razão de verosimilhança

Para amostras com pequena dimensão, o método da matriz de Fisher não é suficientemente conservador. O método da razão de verosimilhança produz resultados que são mais conservadores e é, por conseguinte, mais adequado para estes casos. Para amostras com maior dimensão, não há uma diferença significativa entre os resultados através destes dois métodos (Abernethy, 2006).

O intervalo de confiança pelo método da razão de verosimilhança é calculado pela seguinte equação,

$$-2 \cdot \ln \left(\frac{L(\theta)}{L(\hat{\theta})} \right) \geq \chi^2_{\alpha, k} \quad 4.99$$

onde,

$L(\theta)$ é a função de verosimilhança para o parâmetro desconhecido θ

$L(\hat{\theta})$ é a função de verosimilhança para o parâmetro desconhecido $\hat{\theta}$

$\chi^2_{\alpha, k}$ é a distribuição estatística Qui-quadrado com k graus de liberdade, onde k é o número de quantidades conjuntamente estimadas.

Como os dados da amostra são conhecidos e os parâmetros estimados foram calculados pelo método de máxima verosimilhança, assim, o único termo desconhecido na equação

anterior é o valor da função de verosimilhança para o parâmetro desconhecido, θ . O intervalo de confiança é encontrado com os valores que satisfaçam a equação.

Para as distribuições que têm dois parâmetros, o valor destes dois parâmetros pode variar a fim de satisfazer a equação. Para um dado valor de confiança, α , determina-se uma região de valores conjuntos para θ_1 e θ_2 de modo que a razão da equação seja válida. Esta região pode ser representada graficamente conforme a figura seguinte.

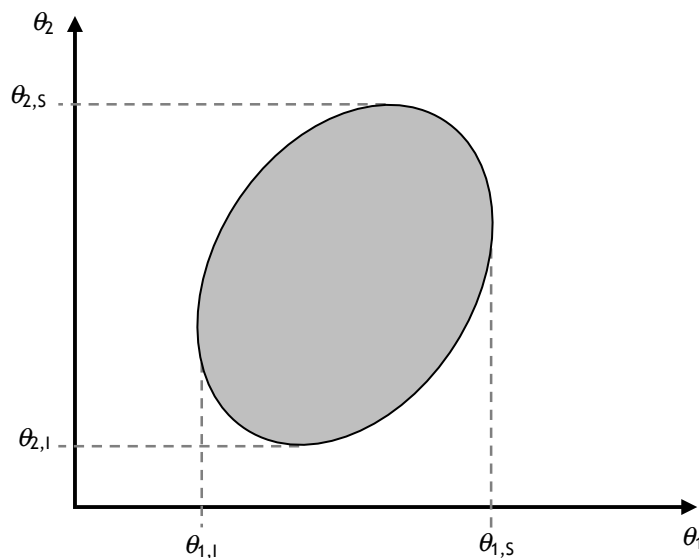


Figura 4.25 - Intervalo de confiança pelo método da razão de verosimilhança.

Os limites dos parâmetros são calculados ao encontrar os valores extremos do contorno da figura anterior em cada eixo para um determinado nível de confiança. Uma vez que cada eixo representa os valores possíveis de um determinado parâmetro, os limites do gráfico de contorno representam os valores extremos dos parâmetros que satisfazem a seguinte equação,

$$-2 \cdot \ln \left(\frac{L(\theta_1, \theta_2)}{L(\hat{\theta}_1, \hat{\theta}_2)} \right) = \chi_{\alpha,1}^2 \quad 4.100$$

Esta equação pode ser reinscrita como,

$$L(\theta_1, \theta_2) = L(\hat{\theta}_1, \hat{\theta}_2) \cdot e^{\frac{-\chi_{\alpha,1}^2}{2}} \quad 4.101$$

O objetivo é encontrar os valores dos parâmetros de modo que a igualdade da equação da razão de verosimilhança anterior seja satisfeita. Não existe uma solução padrão para a resolução, portanto, esses valores devem ser obtidos numericamente.

Uma maneira de resolver a equação é fixar um dos parâmetros e efetuar a iteração do outro até que uma solução aceitável seja alcançada. Isto pode revelar-se bastante complicado.

Em algumas situações é melhor começar os cálculos iterativos com valores próximos dos valores estimados, de modo a assegurar que não se está a realizar cálculos fora da região do gráfico, onde não existe uma solução.

4.3.3 - Método bootstrap

O método bootstrap foi originalmente proposto por Efron em 1979. Embora as ideias subjacentes ao bootstrap sejam mais antigas, o interesse da comunidade científica pelo método surgiu essencialmente após a publicação dos trabalhos de Efron (1979) e Efron e Tibshirani (1986), onde o método bootstrap é apresentado como um método computacional, que tem por base procedimentos de reamostragem para o cálculo de medidas de incerteza dos parâmetros de interesse.

Desde então, o método bootstrap tem vindo a ser aperfeiçoado, constituindo atualmente um método computacionalmente eficaz e amplamente usado na inferência estatística.

O método bootstrap permite o cálculo do intervalo de confiança dos parâmetros de interesse em circunstâncias em que as outras técnicas não são aplicáveis, em particular no caso em que a amostra é pequena ($n < 30$) (Davison, 1997).

Este método considera que a amostra original deve representar a população em estudo. Deste modo, novas amostras são obtidas de forma aleatória a partir de reamostragem da amostra original, denominada de amostra bootstrap, Y_i^* , a qual representa o que seria desejável na prática, se tal fosse possível, repetir a experiência.

Para cada amostra bootstrap é calculada uma estimativa do parâmetro real. No final do processo de reamostragem o conjunto das estimativas obtidas é denominada de estimativa bootstrap, $\hat{\psi}_i^*$.

Para que a aleatoriedade do processo seja minimizada é necessário realizar um grande número de reamostragens, B . O número de reamostragem mínimo é igual a 1000 e para a grande maioria das aplicações, um número de reamostragem igual a 2000 fornece excelentes resultados (Davison, 1997), (Efron, 1993).

Na prática, o número de conjuntos, B , é geralmente selecionado de acordo com um critério de precisão ou de convergência.

O método bootstrap pode ser implementado tanto na forma paramétrica ou na forma não paramétrica. O que difere os dois métodos é a forma de obtenção da amostra.

No caso paramétrico, os parâmetros estimados da distribuição dos dados são conhecidos e a amostra é composta pela realização da amostragem diretamente dessa distribuição.

No caso não paramétrico, a amostra de tamanho n será composta por extrações dos elementos da amostra original com reposição.

4.3.3.1 - Método bootstrap não paramétrico

No método não paramétrico, a reamostragem cria um conjunto de dados, cada um dos quais é replicado a partir da amostra original. Para obter resultados confiáveis, é preciso realizar várias amostras bootstrap do mesmo tamanho n . Essas amostras bootstrap devem ser feitas com reposição e de forma aleatória.

A reamostragem é o nome que se dá a um conjunto de técnicas ou métodos que se baseiam em calcular estimativas a partir de repetidas amostragens dentro da mesma amostra inicial.

Surgiu em meados de 1935, entretanto a aplicação destas técnicas desenvolveu-se mais nos últimos anos, dado que com o avanço tecnológico foi possível utilizar o computador de forma intensiva nos procedimentos de reamostragem.

A amostragem com reposição obtém a amostra a partir da amostra original e volta a colocar a sua informação na amostra original para possivelmente ser usada novamente, ou seja, um ou mais pontos da amostra podem estar ausentes e um ou mais podem ser repetidos mais de uma vez em qualquer conjunto de dados reamostrados.

Uma amostra bootstrap $Y_i^* = (x_1^*, x_2^*, \dots, x_n^*)$ é obtida por amostragem aleatória n vezes, com reposição a partir dos dados originais $X = (x_1, x_2, \dots, x_n)$.

A notação $*$ indica que Y_i^* não é um conjunto de dados reais e sim uma reorganização de $X = (x_1, x_2, \dots, x_n)$. Por exemplo, pode-se obter,

$$x_1^* = x_3, x_2^* = x_5, x_3^* = x_1, \dots, x_n^* = x_6$$

Alguns elementos podem não aparecer e outros aparecer com mais frequência.

Resumidamente, o método de bootstrap gera um grande número de amostras bootstrap B independentes $(Y_1^*, Y_2^*, \dots, Y_B^*)$, por amostragens de tamanho n igual ao da amostra original com reposição da mesma. Correspondente a cada amostra bootstrap tem-se uma réplica bootstrap, que é o valor da estimativa do parâmetro de interesse, denotada por $\hat{\psi}_i^*, i = 1, 2, \dots, B$ (Ibrahim, 2007), (Efron, 1993).

A Figura 4.26 ilustra esse procedimento.

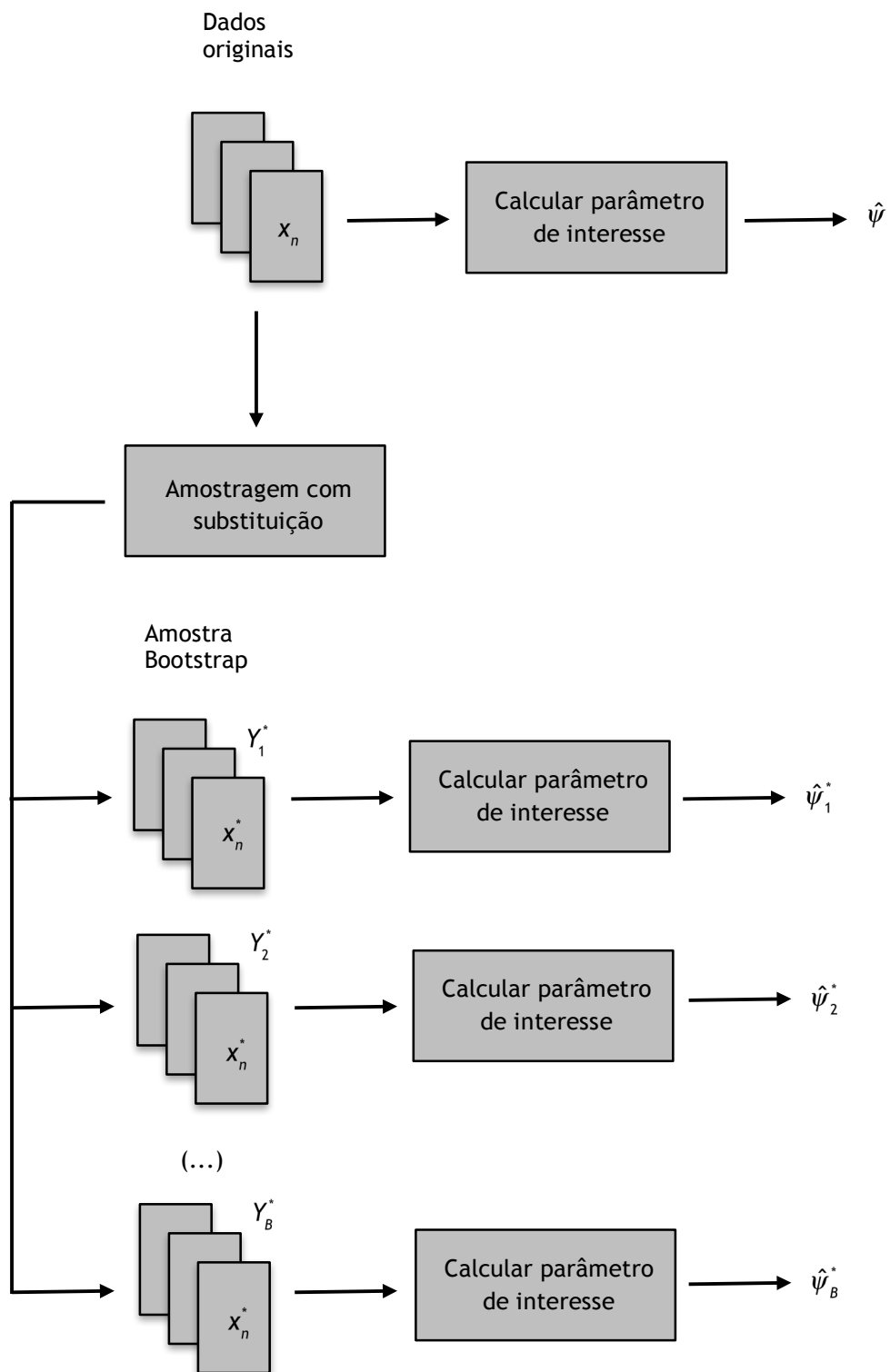


Figura 4.26 - Método bootstrap não paramétrico.

4.3.3.2 - Método bootstrap paramétrico

No caso do bootstrap paramétrico, o método consiste em gerar amostras baseadas na distribuição de probabilidade conhecida utilizando como parâmetros desta distribuição a estimativa dos mesmos, obtida através da amostra geral (Carpenter, 2000).

Dado a amostra $X = (x_1, x_2, \dots, x_n)$ estima-se os parâmetros que caracterizam a distribuição que melhor se ajusta aos dados. Conhecida a distribuição é possível gerar B amostras bootstrap independentes $(Y_1^*, Y_2^*, \dots, Y_B^*)$ de tamanho n .

Do mesmo modo ao método bootstrap não paramétrico, por cada amostra bootstrap tem-se uma réplica bootstrap, que é o valor da estimativa do parâmetro de interesse, denotada por $\hat{\psi}_i^*, i = 1, 2, \dots, B$.

A Figura 4.27 ilustra esse procedimento.

4.3.3.3 - Intervalo de confiança bootstrap

Através deste método é possível construir o intervalo de confiança do parâmetro estimado de uma distribuição, $\hat{\theta}(\hat{\eta}, \hat{\beta})$. Assim o parâmetro de interesse referido anteriormente é dado por $\hat{\psi} = \hat{\theta}$.

Na seção anterior descreve-se como obter amostras bootstrap paramétricas e não paramétricas em geral. Com as amostras bootstrap geradas é possível calcular o desvio padrão das B repetições que será utilizado nos intervalos de confiança, dado por (Hall, 2004), (Efron, 1993),

$$\hat{\sigma}_{boot}(\hat{\theta}^*) = \sqrt{\frac{1}{B-1} \sum_{i=1}^B [\hat{\theta}_i^* - \hat{\theta}^*(.)]^2} \quad 4.102$$

onde,

$$\hat{\theta}^*(.) = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_i^* \quad 4.103$$

A estimativa do enviesamento bootstrap é definida por,

$$\hat{\delta}_{boot}(\hat{\theta}) = \hat{\theta}^*(.) - \hat{\theta} \quad 4.104$$

A seguir apresentam-se os diferentes métodos de obtenção dos intervalos de confiança bootstrap, sendo eles o intervalo bootstrap normal, o intervalo bootstrap-t, o intervalo bootstrap percentil, BCBP e o BCa.

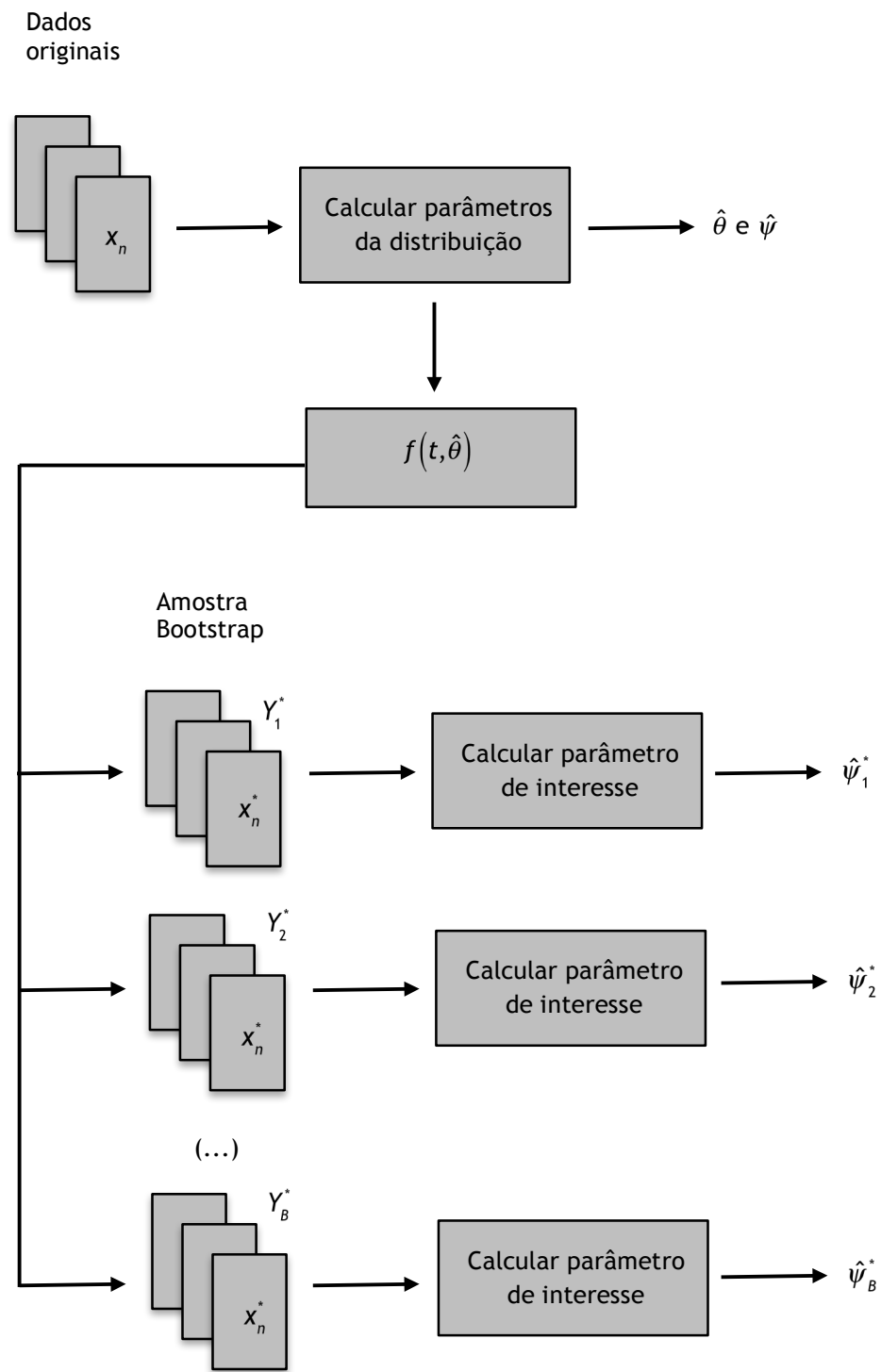


Figura 4.27 - Método bootstrap paramétrico.

4.3.3.3.1 - Intervalo de confiança bootstrap normal

Adotando-se o coeficiente de confiança $100(1 - \alpha)\%$ com $0 < \alpha < 1$ é possível obter o valor z_c na tabela da distribuição normal que satisfaz (Efron, 1993),

$$P[-z_c \leq Z_b^* \leq +z_c] = (1 - \alpha)$$

A variável Z_b^* segue a distribuição normal.

Logo,

$$P\left[-z_c \leq \frac{\hat{\theta} - \theta}{\hat{\sigma}_{boot}(\hat{\theta}^*)} \leq +z_c\right] = (1 - \alpha)$$

Assim, o intervalo de confiança bootstrap normal para θ com coeficiente de confiança 100 (1- α) % é dado por,

$$\left[\hat{\theta} - z_c \cdot \hat{\sigma}_{boot}(\hat{\theta}^*), \hat{\theta} + z_c \cdot \hat{\sigma}_{boot}(\hat{\theta}^*)\right] \quad 4.105$$

Uma vantagem desse método é a facilidade algébrica para obter o intervalo de confiança para θ .

4.3.3.3.2 - Intervalo de confiança bootstrap-t

Da mesma forma que o intervalo de confiança bootstrap normal pode-se encontrar o intervalo de confiança bootstrap-t (Fang, 2015).

Ao fixar-se o coeficiente de confiança em 100 (1- α) % com $0 < \alpha < 1$, obtém-se na tabela da distribuição t-Student o valor t_c que satisfaz,

$$P[-t_c \leq T_b^* \leq +t_c] = (1 - \alpha)$$

A variável T_b^* segue a distribuição t-Student.

Logo,

$$P\left[-t_c \leq \frac{\hat{\theta} - \theta}{\hat{\sigma}_{boot}(\hat{\theta}^*)} \leq +t_c\right] = (1 - \alpha)$$

Assim, o intervalo de confiança bootstrap-t para θ com coeficiente de confiança 100 (1- α) % é dado por,

$$\left[\hat{\theta} - t_c \cdot \hat{\sigma}_{boot}(\hat{\theta}^*), \hat{\theta} + t_c \cdot \hat{\sigma}_{boot}(\hat{\theta}^*)\right] \quad 4.106$$

Há uma tendência geral para que os intervalos bootstrap-t tenham amplitudes menores do que os intervalos baseados na tabela normal.

4.3.3.3.3 - Intervalo de confiança bootstrap percentil

Este modo de determinar o intervalo de confiança baseia-se na utilização do conceito estatístico de percentil para definir os limites de confiança. O intervalo de confiança bootstrap percentil com coeficiente de confiança $100(1-\alpha)\%$ é obtido pelos $(\alpha/2)^{ésimo}$ e $(1-\alpha/2)^{ésimo}$ percentil do histograma bootstrap (Arasan, 2014), (Meeker, 1998).

Assim,

$$[L_i, L_s] = \left[\hat{\theta}_{\left(\frac{\alpha}{2}\right)}^*, \hat{\theta}_{\left(1-\frac{\alpha}{2}\right)}^* \right] \quad 4.107$$

A distribuição bootstrap da média de $\hat{\theta}$ aproxima-se da distribuição normal quando $n \rightarrow \infty$. Nesse caso, os intervalos de confiança bootstrap percentil e o intervalo de confiança bootstrap normal são semelhantes. Quando o n é reduzido, o histograma bootstrap pode afastar-se da normalidade e os dois tipos de intervalos podem ser diferentes.

A simplicidade é a atração deste método, o que explica a sua contínua popularidade. Ao contrário do bootstrap-t, não é necessário calcular a estimativa do desvio padrão das B réplicas.

Quando a assimetria está bastante presente é mais recomendável que se utilize o método BCPB e o método BCa (Efron, 1993).

4.3.3.3.4 - Intervalo de confiança bootstrap BCPB

O intervalo de Confiança bootstrap BCPB “Biased-Corrected Percentilt bootstrap”, é um método que faz correções substanciais. Neste método os extremos do intervalo são os percentis da distribuição bootstrap ajustados, para corrigir o enviesamento e a assimetria da distribuição (Efron, 1993).

Considere-se as B réplicas bootstrap $\hat{\theta}_i^*$, $i = 1, 2, \dots, B$, ordenadas de forma crescente. Assim encontra-se a proporção das réplicas bootstrap menores que $\hat{\theta}$, denominada por p_0 .

A proporção p_0 é definida como a probabilidade de uma estimativa ser inferior à estimativa da amostra original $\hat{\theta}$.

A expressão para a proporção p_0 das réplicas bootstrap é dada por,

$$p_0 = P[\hat{\theta}_i^* \leq \hat{\theta}], \quad i = 1, 2, \dots, B \quad 4.108$$

Com o valor de p_0 determina-se o parâmetro de correção do enviesamento que é dado por z_0 .

$$z_0 = \Phi^{-1}(p_0) \quad 4.109$$

Onde Φ^{-1} é a inversa da função de distribuição acumulada normal.

Assim tem-se B valores de z_0 e em seguida calcula-se a média para esses valores obtendo-se um único valor, designado por \bar{z}_0 .

Seleciona-se um coeficiente de confiança 100 (1- α) % para a estimativa do parâmetro e encontrado $z_{(\alpha/2)}$ obtém-se as correções para os percentis.

As correções P_l e P_s são definidas por,

$$P_l = \Phi \left(2\bar{z}_0 - z_{\left(\frac{\alpha}{2}\right)} \right), \quad P_s = \Phi \left(2\bar{z}_0 + z_{\left(\frac{\alpha}{2}\right)} \right) \quad 4.110$$

O intervalo de confiança bootstrap BCPB é dado por,

$$[L_l, L_s] = \left(\hat{\theta}_{(P_l)}^*, \hat{\theta}_{(P_s)}^* \right) \quad 4.111$$

4.3.3.3.5 - Intervalo de confiança bootstrap BCa

O método de Correção do Enviesamento Acelerado BCa (“Bias-Corrected and Acceleration”) permite encontrar o intervalo de confiança quando a assimetria estiver bastante presente. Esta metodologia de construção de intervalos de confiança constitui-se como um aperfeiçoamento na conceção dos intervalos de confiança bootstrap (Balakrishnan, 2007), (Fang, 2015).

No entanto, os intervalos de confiança BCa exigem um volume de cálculo elevado, fator que se constitui como um grande inconveniente para este método.

Assim como nos casos dos intervalos de confiança percentil e BCPB, o BCa utiliza os percentis da distribuição bootstrap para a construção dos intervalos de confiança para os parâmetros de interesse. Neste caso utiliza-se percentis que dependem de duas constantes, \hat{z}_0 designado por correção para a tendência e \hat{a} designado de constante de aceleração, que ajusta o intervalo de confiança em relação a assimetria.

A constante \hat{z}_0 representa a discrepância entre a mediana de $\hat{\theta}^*$ e $\hat{\theta}$ e é calculado pela seguinte equação,

$$\hat{z}_0 = \Phi^{-1} \left(\frac{n^\circ \{ \hat{\theta}_i^* < \hat{\theta} \}}{B} \right) \quad 4.112$$

Onde Φ^{-1} é a inversa da função de distribuição acumulada normal.

A constante de aceleração \hat{a} representa a taxa de variação do desvio padrão de $\hat{\theta}$ relativamente ao verdadeiro valor do parâmetro θ e é dada pela seguinte equação,

$$\hat{a} = \frac{\sum_{i=1}^n (\hat{\theta}_{(.)} - \hat{\theta}_{(i)})^3}{6 \left[\sum_{i=1}^n (\hat{\theta}_{(.)} - \hat{\theta}_{(i)})^2 \right]^{\frac{3}{2}}} \quad 4.113$$

onde,

$$\hat{\theta}_{(.)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)} \text{ e}$$

$\hat{\theta}_{(i)}$ representa uma amostra Jackknife (após eliminar a observação i).

O intervalo de confiança bootstrap BCa com coeficiente de confiança de 100 $(1-\alpha)$ % é dado por,

$$[L_l, L_s] = (\hat{\theta}_{(p_l)}^*, \hat{\theta}_{(p_s)}^*)$$

onde,

$$P_l = \phi \left(\frac{\hat{z}_0 + z_{\left(\frac{\alpha}{2}\right)}}{1 - \hat{a} \left(\frac{\hat{z}_0 + z_{\left(\frac{\alpha}{2}\right)}}{\left(\frac{\alpha}{2}\right)} \right)} \right), \quad 4.114$$

e

$$P_s = \phi \left(\frac{\hat{z}_0 + z_{\left(\frac{\alpha}{2}\right)}}{1 - \hat{a} \left(\frac{\hat{z}_0 + z_{\left(1-\frac{\alpha}{2}\right)}}{\left(1-\frac{\alpha}{2}\right)} \right)} \right) \quad 4.115$$

sendo $\phi(\cdot)$ a função de distribuição normal e $Z_{(\alpha/2)}$ o percentil de ordem 100. $Z_{(\alpha/2)}$ da distribuição normal.

Se o estimador do parâmetro de interesse tiver uma distribuição aproximadamente normal e tiver um enviesamento pequeno, os intervalos de confiança bootstrap-t e o bootstrap percentil são adequados e os valores dos limites de confiança são próximos. Caso os intervalos de confiança de bootstrap-t e o bootstrap percentil não apresentem valores próximos, os métodos não são adequados, ou seja, existem métodos melhores.

Se a distribuição do estimador de interesse não for aproximadamente normal e/ou apresentar um enviesamento muito grande, os métodos BCPB e BCa são adequados. Esses

métodos também são adequados quando a distribuição apresentar assimetria de maneira acentuada.

4.4 - Conclusões

Com a descrição e comparação dos métodos para a estimação dos parâmetros da distribuição de Weibull apresentada neste capítulo, estão reunidas as condições necessárias para a escolha do método que melhor se adequa perante diferentes cenários, nomeadamente, para o caso prático apresentado no capítulo seguinte.

O método de máxima verosimilhança apresenta melhores resultados para dados completos e censurados à direita, do que o método dos mínimos quadrados. Contudo, quando a precisão dos valores obtidos não é um elemento fundamental, o método dos mínimos quadrados, conjuntamente com a representação gráfica, apresenta-se como uma boa alternativa, devido à sua implementação simples, com relativa precisão dos valores obtidos e permite uma visualização gráfica do comportamento dos dados.

Recomenda-se a utilização do método de Bernard para dados completos e do método de Herd-Johnson para dados censurados na estimação de $F(t)$.

Neste capítulo, foi possível também verificar o desenvolvimento do algoritmo EM como método numérico para determinar as soluções da equação que deriva do método da máxima verosimilhança, em particular, no caso da presença de dados censurados à direita.

Capítulo 5

Aplicação experimental

5.1 - Introdução

Um dos objetivos deste capítulo é a aplicação do algoritmo EM apresentado no capítulo 4 para determinar as soluções da função obtida pelo método de máxima verosimilhança na estimação dos parâmetros da distribuição de Weibull, a partir de dados censurados à direita provenientes do registo histórico de falhas do componente em análise.

Neste capítulo é descrito e aplicado um conjunto de procedimentos que permite identificar quais os componentes mais críticos do sistema e assim incidir especificamente sobre estes componentes a aplicabilidade do algoritmo.

A estimação dos parâmetros da distribuição também é realizada pelo método dos mínimos quadrados.

Para o método de máxima verosimilhança é apresentado o intervalo de confiança pelo método bootstrap dos parâmetros estimados da distribuição de Weibull.

5.2 - Bomba centrífuga

O sistema em análise é um grupo de bombas centrífugas da refinaria Galp Energia, em Matosinhos.

A bomba centrífuga é uma das máquinas elétricas mais utilizada na indústria petroquímica. Este tipo de bomba possui uma posição cimeira no transporte de fluidos. Estima-se que entre todas as bombas instaladas, na indústria petroquímica, entre 80 a 90%, são bombas centrífugas (Bloch, 2010). Devido às suas excelentes características, a bomba centrífuga é utilizada em diversas aplicações, desde o bombeamento de água até ao transporte de fluidos inflamáveis a elevada pressão e temperatura (Di Febo, 2015), (Girdhar, 2005).

Por ser um equipamento muito importante no processo industrial, pode ser instalada com uma bomba de reserva, de modo a garantir a continuidade operacional em caso de avaria da bomba principal.

As bombas hidráulicas industriais apresentam diversas características que permitem diferenciá-las e classificá-las sob diversos aspetos.

A classificação mais comum divide as bombas hidráulicas em dois grandes grupos, em bombas dinâmicas ou turbo-bombas e bombas volumétricas ou de deslocamento positivo (Bloch, 2010), (Girdhar, 2010), (Palgrave, 2003). As bombas centrífugas integram o primeiro grupo.

Nas turbo-bombas também designadas por bombas dinâmicas, é transmitido ao fluido uma aceleração para que esta adquira energia cinética a partir da transformação da energia mecânica por meio do movimento do rotor inserido no corpo da bomba.

De acordo com as diferentes formas e tipos de rotor, as turbo-bombas, podem ser classificadas como (Forsthoffer, 2005):

- Bombas radiais ou centrífugas: a trajetória do fluido faz-se segundo um plano radial (normal ao eixo), do centro para a periferia do rotor.
- Bombas mistas ou diagonais: a trajetória do fluido é diagonal relativamente ao eixo, assim, situa-se entre as bombas centrífugas e as bombas axiais.
- Bombas axiais: a trajetória do fluido é segundo a direção do eixo da bomba, são mais utilizadas para aplicações onde é necessário um grande caudal e baixa altura manométrica.

As turbo-bombas têm um processo de fabrico mais económico do que as bombas volumétricas, são mais compactas e adequam-se para uma grande gama de alturas de elevação e caudais.

Nas bombas volumétricas ou de deslocamento positivo, a transferência de energia é feita por variações de volume que ocorrem enquanto o fluido está confinado numa câmara.

O movimento do fluido é causado pela ação do órgão de impulsão da bomba que obriga o fluido a executar o mesmo movimento a que está sujeito este impulsor (êmbolo, engrenagens, lóbulos, palhetas), em movimentos alternados. Dá-se o nome de volumétrica porque o fluido ocupa e sai, de forma sucessiva espaços no interior da bomba, com volumes conhecidos. A designação de deslocamento positivo deve-se ao facto do movimento geral deste fluido se dar na mesma direção das forças transmitidas.

O deslocamento positivo do cursor da bomba define o volume bombeado por cada ciclo de operação.

As bombas volumétricas podem ser classificadas em:

- Bombas rotativas: comandada por um movimento de rotação (engrenagens, lóbulos, palhetas, fusos ou parafusos).

- Bombas de êmbolo ou alternativas: o órgão que produz o movimento do fluido é um pistão que, em movimentos alternativos, aspira e expulsa o fluido bombeado (pistão, diafragma, membrana).

As bombas volumétricas são mais adequadas para pressões elevadas e caudais baixos.

A figura 5.1 apresenta a classificação das bombas hidráulicas.

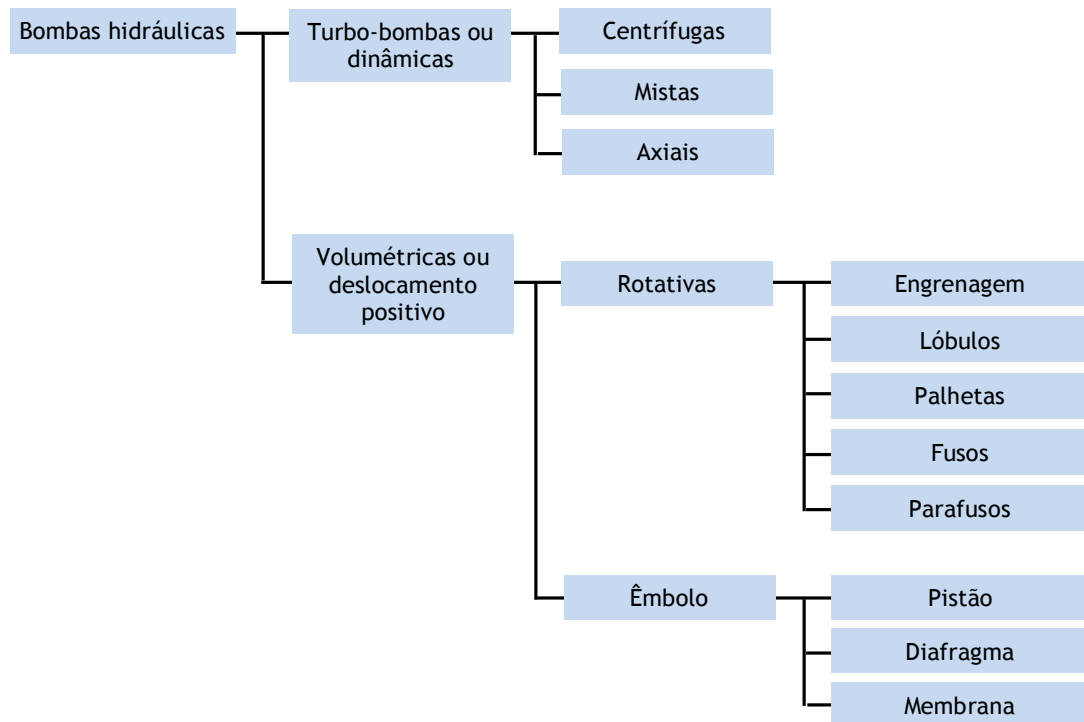


Figura 5.1 - Classificação das bombas hidráulicas (Bloch, 2010).

A bomba centrífuga é amplamente utilizada porque tem um princípio de funcionamento simples, que se encontra materializado com aspetos construtivos de fácil execução, apresenta uma elevada eficiência e facilidade de operação e se for bem aplicado, permite obter sistemas estáveis e com boa qualidade (Girdhar, 2005).

As bombas são máquinas que transferem energia ao fluido com a finalidade de transportá-lo obedecendo às condições do processo. Recebem energia de uma fonte externa, neste estudo em particular um motor eléctrico, e cedem essa energia ao fluido sob a forma de pressão, energia cinética, ou ambas, isto é, aumentam a pressão e/ou a velocidade do fluido.

O movimento do fluido ocorre por meio de ação de forças que se desenvolvem através da rotação de um eixo acoplado à roda (rotor, impulsor) dotado de pás, na qual recebe o fluido e o conduz pela periferia, sob ação da força centrífuga.

Estes equipamentos podem ser constituídos por um único ou vários rotores, dentro do corpo (carcaça), assentes sobre o mesmo eixo. No primeiro caso são denominadas de simples estágio, no outro caso de múltiplos estágios.

Em relação a estas últimas, são adequadas para sistemas que precisam de atingir caudais ou pressões elevadas, sendo o efeito semelhante ao da colocação de bombas em série.

5.3 - Caracterização do caso de estudo

Conforme referido anteriormente, as bombas centrífugas presentes numa refinaria petrolífera são bastante extensas, com características e aplicações diferenciadas. Assim foi escolhido um conjunto de bombas centrífugas com base nos seguintes critérios:

- Características e aplicação semelhante;
- Tempo de operação de cada bomba;
- Histórico de manutenção bem documentado.

A metodologia experimental correspondeu ao levantamento do histórico de falhas de cinco bombas centrífugas, utilizadas no transporte de óleo com densidade semelhante, referente ao período de 2006 a 2013.

Todas as informações relativas às ações de manutenção das bombas centrífugas foram obtidas através dos relatórios de manutenção registados no software de gestão SAP da empresa.

Por estarem inseridas num processo produtivo crítico, as bombas em estudo estão instaladas com uma bomba de reserva, a fim de garantir a continuidade operacional em casos de falha da bomba principal.

Outro aspeto relevante nesse estudo, foi limitar a idade máxima das bombas centrífugas em 15 anos, para assim poder apresentar resultados das bombas mais atuais e evitar resultados deturpados das bombas mais antigas que normalmente apresentam elevadas taxas de falha.

Todas as bombas da refinaria encontram-se ao abrigo do contrato de inspeção periódica. A inspeção efetuada contempla análise de vibrações, ruído e análise visual.

De acordo com o definido pela área de inspeção dinâmica da refinaria (que gere o contrato de inspeção), a periodicidade dos vários tipos de inspeções poderá ser mensal, semanal ou diária. Contudo, o operador da unidade local observa todos os seus equipamentos pelo menos uma vez por turno (com registo de observações).

O que implica que as bombas centrífugas em análise são visualmente inspecionadas pelo utilizador do equipamento, pelo menos a cada 8 horas.

Na figura 5.2 são indicados os elementos constituintes da bomba centrífuga analisada neste estudo.

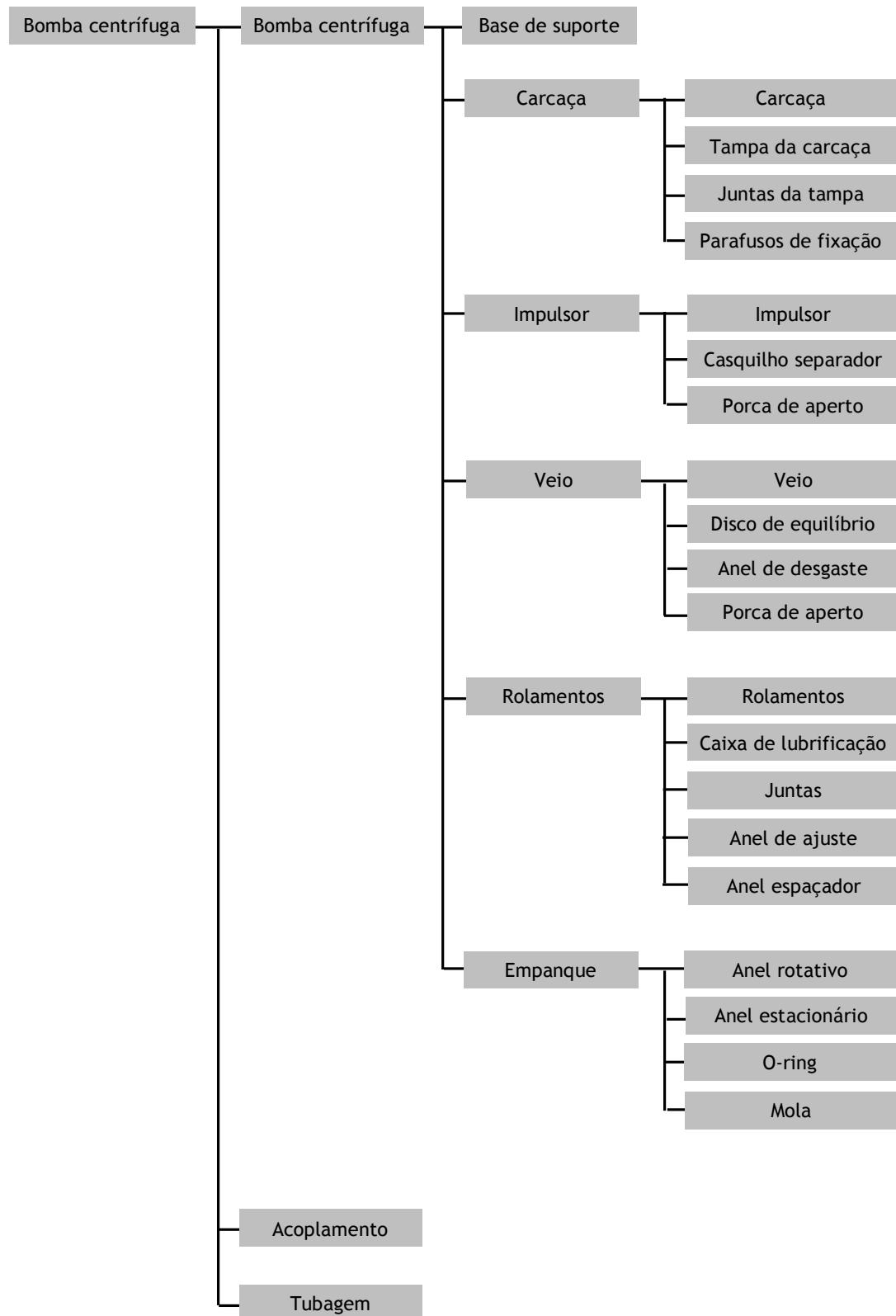


Figura 5.2 - Árvore funcional da bomba centrífuga.

As bombas centrífugas em análise têm o acionamento através de um motor elétrico. Neste estudo o motor elétrico não é considerado no sistema. O limite do sistema é no acoplamento entre a bomba e o motor elétrico.

Este procedimento vai de encontro ao estudo apresentado pela Oreda (Offshore Reliability Data Handbook) (Oreda, 2002).

Pode-se verificar na Tabela 5.1, que dois tipos de componentes sobressaem pelo seu número de falhas, que são o empanque mecânico e os rolamentos, com 44 e 32 falhas, respetivamente, no total de 97 falhas. Assim os empanques e os rolamentos representam 45,4% e 33,0%, respetivamente, do número total das falhas.

A mesma tabela apresenta a comparação com os resultados apresentados na referência bibliográfica (Bloch, 2010). Os dados recolhidos nesse estudo são referentes a falhas de 3300 bombas durante o período de 2003 e 2004, de empresas petroquímicas.

Tabela 5.1 - Número e percentagem de falhas, por componente e comparação com (Bloch, 2010).

| Componentes | Sistema em análise | |
|-------------------|--------------------|------|
| | nº | % |
| Empanque mecânico | 44 | 45,4 |
| Rolamentos | 32 | 33,0 |
| Impulsor | 11 | 11,3 |
| Base de suporte | 5 | 5,2 |
| Veio | 3 | 3,1 |
| Carcaça | 2 | 2,1 |
| Total | 97 | 100 |

| Componentes | (Bloch, 2010) |
|-------------------|---------------|
| | % |
| Empanque mecânico | 43 |
| Rolamentos | 40 |
| Impulsor | 12 |
| Base de suporte | 2 |
| Acoplamento | 2 |
| Outros | 1 |
| Total | 100 |

Existem diferenças nos dois estudos, o que torna a comparação direta dos resultados difícil. No entanto, é possível encontrar muitos aspetos em comum entre os dois estudos.

Ambos os estudos evidenciam que, em conjunto, as falhas relativas aos empanques e aos rolamentos, correspondem aproximadamente a 80% do número total de falhas registado, conforme se pode verificar na figura 5.3.

Por esta razão, a fiabilidade do empanque e dos rolamentos, nestes tipos de bombas, é um aspeto fundamental para o sistema, na manutenção, na segurança e nos níveis de desempenho elevados.

Assim, a partir da apreciação destes resultados, justifica-se a importância de um estudo mais pormenorizado nestas áreas.

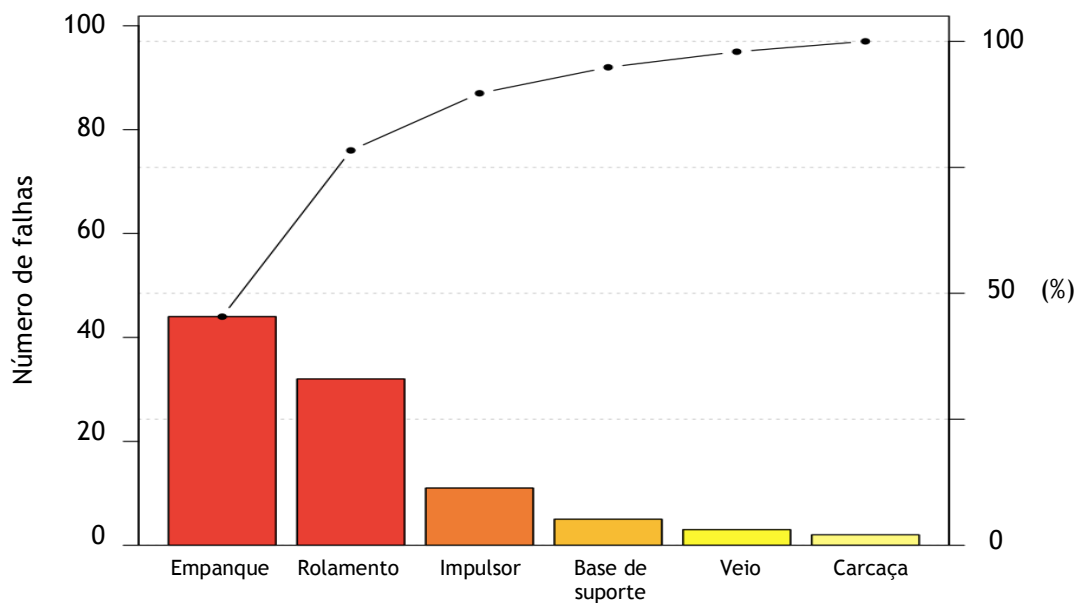


Figura 5.3 - Gráfico de Pareto das bombas centrífugas por componente.

5.3.1 - Empanque mecânico

Conforme indicado anteriormente, numa bomba centrífuga, o empanque mecânico é o componente responsável por aproximadamente 45% das falhas.

O empanque mecânico tem a função de impedir a fuga de fluido, em veios rotativos, para o exterior da bomba.

Existem diferentes modelos para atender a aplicações específicas.

Normalmente os empaques mecânicos são compostos por uma parte fixa (estacionário) que permanece estática e por uma parte rotativa que roda de forma solidária com o veio. As faces rotativa e estacionária encontram-se perpendiculares ao eixo (Bachus, 2003), (Forsthoffer, 2005), (Mobley, 1999) conforme apresentado na figura 5.4.

Nestes conjuntos devem sempre existir conjuntamente as suas respetivas vedações secundárias (o-ring).

A parte rotativa tem uma ou mais molas, cuja função é manter o contacto entre a superfície desta e a superfície da parte fixa.

Os empanques com múltiplas molas geralmente são mais caros e mais difíceis de montar, devido ao maior número de componentes. Quanto maior for a mola de um empanque com uma única mola maior pode ser a resistência à corrosão, devido às maiores dimensões da mola, embora as molas utilizadas em empanques mecânicos mais recentes são bastante resistentes a quase qualquer líquido presente na indústria. Múltiplas molas também resultam numa distribuição de carga mais uniforme. A maioria dos empanques utilizados nas indústrias utiliza múltiplas molas.

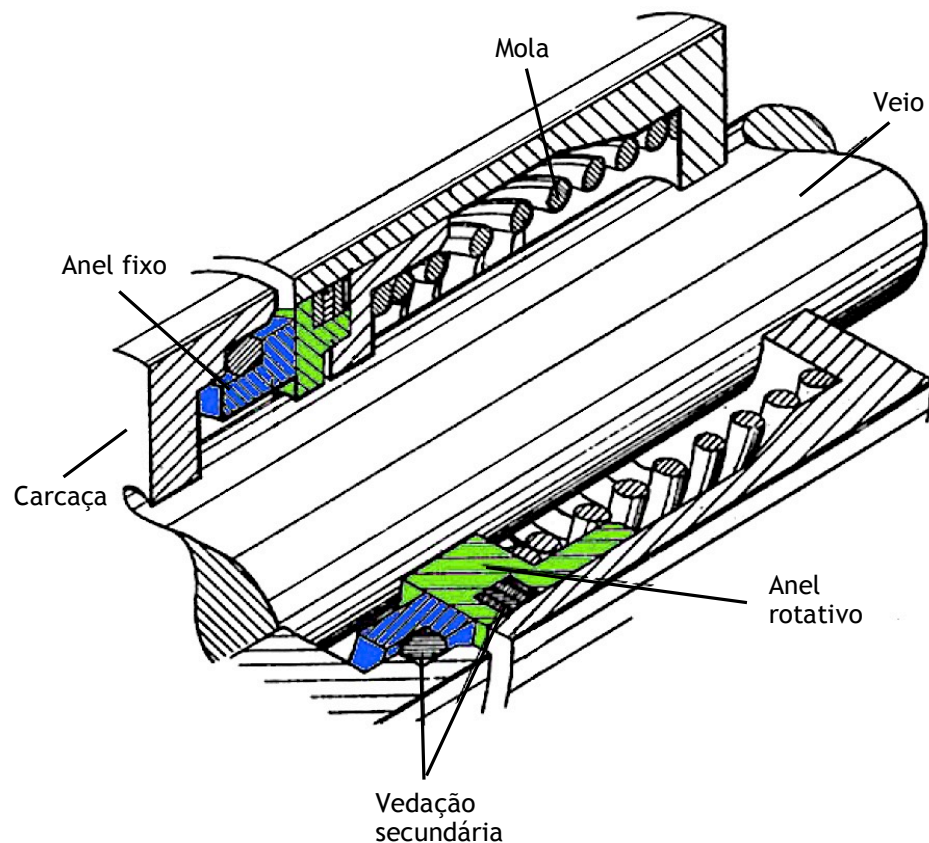


Figura 5.4 - Empanque mecânico de uma bomba centrífuga (Forsthoffer, 2005).

O contato axial estabelecido pela força exercida pela mola e pela pressão do fluido determina o fecho entre as faces e impede a passagem do fluido para o eixo da bomba, normalmente designa-se por vedação primária ou dinâmica.

Já a vedação secundária ou estática é feita normalmente por anéis (o-ring) ou foles de metal, impedindo a passagem do fluido pelos interstícios do empanque mecânico.

Os o-rings são facilmente disponíveis em vários tamanhos e materiais, normalmente são baratos e podem estar sujeitos a uma pressão mais elevada, quando comparado com o fole de metal. A principal desvantagem do o-ring é a limitada faixa de temperatura permitida. A alta temperatura, tipicamente, é a razão pela qual os empanques com foles metálicos são utilizados. A introdução de compostos mais recentes permitiu a utilização de o-rings a temperaturas mais altas.

A maioria dos empanques mecânicos utilizados na indústria utiliza o-rings como elemento de vedação secundário. A vedação secundária com elastômeros é normalmente utilizada apenas para temperaturas até 200 ou 250 °C.

Os materiais utilizados na produção dos empanques mecânicos são muito variados e determinam-se pelo fluido que possa estar em contacto com o empanque.

O desgaste entre as superfícies em contacto ocorre apesar de existir uma película de fluido entre as superfícies para eliminar o máximo possível a fricção.

Para a maioria dos empanques mecânicos os dois pontos mais críticos de contato ocorrem entre os seguintes:

- Entre o anel rotativo e o anel estacionário;
- Entre o anel rotativo e o veio.

O funcionamento do empanque mecânico é o resultado da interação entre o atrito de superfícies sólidas e as forças hidrodinâmicas geradas pelo fluido na interface devido ao movimento relativo das faces. A combinação destes dois efeitos gera forças que mantêm as duas faces ligeiramente afastadas quando em operação.

A formação da película líquida entre as superfícies é semelhante ao fenómeno hidrodinâmico que ocorre num rolamento, exceto que a carga axial geralmente é muito menor.

O fluido é frequentemente um lubrificante pobre, as temperaturas podem ser muito elevadas, e a película de fluido nem sempre pode separar as faces completamente e muitas vezes são parcialmente vaporizados, como no caso de gás liquefeito de petróleo ou de água quente.

Quando um empanque mecânico começa a rodar sob pressão, as superfícies do empanque ficam sujeitas à força da mola, à força de pressão estática do fluido e à força hidrodinâmica gerada pelo fluido. Um dos fatores que influênciam a vida útil do empanque é a fração da pressão total exercida pela pressão hidrodinâmica do fluido.

Se a espessura do filme é suficientemente grande para evitar completamente o contacto sólido, o desgaste é muito baixo. Ao contrário de um rolamento hidrodinâmico, o empanque mecânico, normalmente, não é projetado para funcionar com um filme líquido espesso, pois isso levaria a fuga elevada. Este é o dilema do empanque mecânico.

As condições de lubrificação que resultam numa vida útil mais longa também resultam em mais fugas. Estes, claramente, são requisitos contraditórios.

A necessidade de poucas fugas de fluido significa que a espessura da película deve ser muito fina. Para obter boas condições de lubrificação com uma película fina, a rugosidade da superfície deve ser muito pequena. As superfícies do empanque normalmente têm um bom acabamento superficial para reduzir o contato.

Mesmo quando as superfícies têm um acabamento superficial reduzido, as faces do empanque têm sempre alguma rugosidade. Em condições normais de operação ocorre algum contato sólido. Este contacto sólido provoca a remoção de pontos mais altos da superfície.

Por vezes, são utilizadas configurações especiais para melhorar a lubrificação, quando o fluido é um lubrificante pobre.

O empanque mecânico mesmo em bom estado permite sempre uma passagem desprezível de fluido, que é vantajoso pelos motivos referidos. A falha ocorre quando esta fuga aumenta para níveis inaceitáveis. A fuga admissível de fluido pelo empanque mecânico pode depender de muitos fatores, incluindo a toxicidade e inflamabilidade do fluido, e dos regulamentos existentes.

A figura seguinte ilustra a interação entre o contacto de superfície e a fuga normal de fluido.

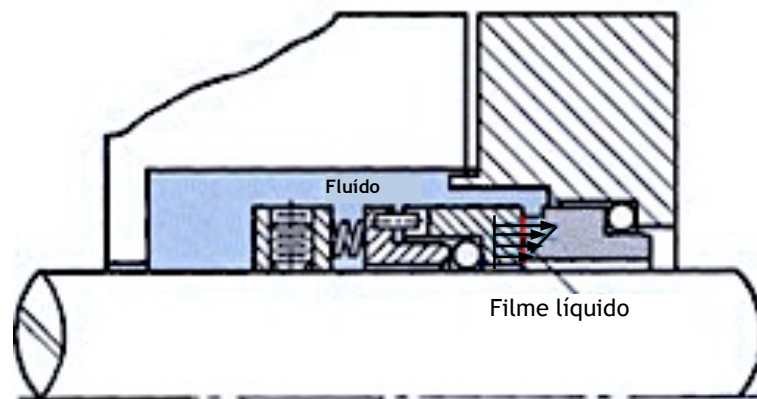


Figura 5.5 - Representação das forças envolvidas num empanque mecânico (Bloch, 2010).

Os modos de falha mais comuns são os seguintes (Bachus, 2003), (Bloch, 2010), (Karassik, 2001):

- Excessiva fuga de fluido;
- Baixa resposta mecânica.

As causas das falhas mais comuns podem ser devidos a:

- **Condições de funcionamento:** as condições de funcionamento, designadamente, a temperatura, a pressão, a viscosidade, a pureza do fluido (contaminação) e o meio envolvente, influenciam a longevidade do empanque mecânico conforme referido anteriormente.

As partículas abrasivas presentes no fluido durante a operação têm uma forte influência sobre a resistência ao desgaste do empanque. A taxa de desgaste do empanque aumenta com a quantidade de contaminação. Os contaminantes do meio ambiente, como por exemplo, a areia pode entrar no sistema do fluido e causar danos generalizados no empanque mecânico nomeadamente nas juntas elastoméricas e nas superfícies de contacto do empanque.

Outro exemplo, é relativo ao aumento de pressão do fluido que pode criar calor extra na face de contacto, que por sua vez pode aumentar o desgaste e o surgimento de outros modos de falha destrutivos tal como a fratura. O aumento da pressão também influencia o estado das juntas elastoméricas (o-ring) devido à ação de extrusão e por consequência o seu dano. A figura seguinte ilustra a ação de extrusão no o-ring.

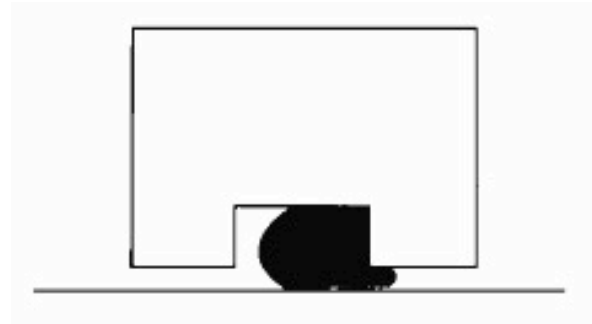


Figura 5.6 - Representação da ação de extrusão no o-ring (Bachus, 2003).

Conforme mencionado anteriormente, a operação a alta temperatura pode danificar o empanque mecânico.

Na vedação secundária feita com elastômeros, o limite de temperatura geralmente não é muito alta. A maior parte dos elastômeros utilizados na indústria ficam danificados, se forem expostos a temperaturas superiores a 200 ou 250 °C. Um elastômero superaquecido normalmente torna-se duro e não é capaz de se deformar elasticamente, como inicialmente previsto. Essa incapacidade para se adaptar ao movimento de outras partes resulta em fugas.

Em regra, o aumento da temperatura e da pressão aumenta a probabilidade de ocorrência de falhas no empanque.

- **Erro humano:** geralmente por não respeitarem a regulamentação existente ou por falta de conhecimento como, por exemplo, a deficiente instalação, o projeto errado ou a operação inadequada.
- **Corrosão:** o mecanismo de corrosão dos vários componentes do empanque depende do componente específico a ser analisado. Ao contrário da corrosão do metal, que é quase sempre um processo eletroquímico, os anéis de elastômeros podem ser sujeitos a processos de corrosão específicas.

Os elementos metálicos do empanque mecânico devem ser completamente inertes ao fluido selado. O menor índice de corrosão nas faces do empanque prejudica a capacidade de vedação. Na maioria das vezes, o resultado da corrosão surge de uma aplicação incorreta ou da instalação de um componente diferente do originalmente especificado.

Os materiais habitualmente utilizados na vedação secundária são geralmente inertes para quase todos os fluidos encontrados na indústria petroquímica. Mais uma vez, os

danos do elastômero (o-ring) muitas vezes é o resultado de uma má aplicação. Os elastômeros podem inchar se houver a absorção do fluido. Este inchaço resulta num aumento de diâmetro, o que pode prejudicar a circulação do empanque ou mesmo fraturar algum componente. O aumento de volume admissível para um anel de vedação estático é de cerca de 50% em volume. O o-ring dinâmico pode começar a não funcionar corretamente se o aumento do tamanho for superior a 15 ou 20% em volume. Como mencionado anteriormente, o aumento do volume, normalmente, é devido a má aplicação.

As causas de falha mencionadas anteriormente, desenvolvem mecanismos de falha essencialmente mecânicas, como por exemplo, desgaste, fratura ou deformação.

Os mecanismos de falha mais habituais são o desgaste e a perda das propriedades do o-ring, conforme referido anteriormente, contudo outros mecanismos podem surgir, como por exemplo a fratura.

A fratura do empanque pode acontecer devido a vários fatores.

Algumas causas são mais óbvias, como por exemplo, a instalação incorreta, outras são menos óbvias, mas que também podem surgir, como é o caso da fratura devido ao choque térmico. As tensões internas que se desenvolvem a partir de gradientes térmicos podem levar à fratura do anel do empanque mecânico. Excessiva vibração e binário são outras das causas que podem provocar o mesmo efeito.

Na tabela 5.2, são indicados os modos mais comuns de falha, respetivos mecanismos e causas no empanque mecânico de uma bomba centrífuga (Bachus, 2003), (Bloch, 2010), (Karassik, 2001).

Para o desenvolvimento do estudo foi necessário assumir determinadas considerações. Os empanques mecânicos são considerados essencialmente como componentes não reparáveis.

Para os propósitos deste estudo, selecionamos apenas os dados de falha devido à fuga excessiva de fluido para o exterior. A razão para esta consideração é devido ao seu número significativo de falhas e para restringir o estudo a apenas um modo de falha.

As falhas consecutivas são tratadas como dados completos. As bombas centrífugas em análise são visualmente inspecionadas pelo utilizador do equipamento pelo menos a cada 8 horas; assim comparativamente ao tempo total em análise desprezou-se o possível erro entre duas inspeções e assumiu-se que se conhece o momento exato de falha e não se considera os dados censurados por intervalo.

O último registo em cada bomba não corresponde a uma falha, mas ao fim do ensaio, dado que as bombas continuaram em funcionamento. Assim o último registo de tempo em cada bomba foi considerado como dado censurado à direita.

Com estas considerações é possível caracterizar o componente em análise e determinar os parâmetros da distribuição estatística que melhor se ajusta aos tempos de falha do empanque mecânico através da metodologia apresentada no capítulo 2.

Tabela 5.2 - Modos, causas e respectivos mecanismos de falha no empanque mecânico.

| Modo de falha | Mecanismo de falha | Causa da falha |
|-------------------------|----------------------|--|
| Excessiva fuga | Desgaste | Desalinhamento |
| | | Excessivo binário |
| | | Inadequada lubrificação |
| | | Empeno do veio |
| | | Excessiva pressão |
| | | Pobre acabamento superficial |
| | | Contaminação |
| | Rutura da mola | Defeito do material |
| | | Corrosão |
| | | Desalinhamento |
| | Fratura | Excessiva pressão |
| | | Corrosão sob tensão |
| | | Instalação incorreta |
| | | Excessivo binário |
| | | Inadequada lubrificação |
| | | Vibração excessiva |
| | | Choque térmico |
| | Fragilização | Contaminação |
| | | Incompatibilidade entre o fluido e o empanque |
| | | Degradação térmica |
| | | Longos períodos de inatividade entre utilizações |
| | Deformação do o-ring | Corrosão |
| | | Excessiva pressão |
| | | Elevada temperatura |
| | | Erro de montagem |
| | | Inadequada lubrificação |
| | Deformação da face | Gradiente térmico |
| | | Excessiva pressão |
| | | Desalinhamento |
| | | Partículas presas na face |
| Baixa resposta mecânica | Excessiva fricção | Excessiva pressão |
| | | Contaminação |
| | | Cristalização do fluido |
| | | Desalinhamento |
| | | Empeno do veio |

Nesta fase, os passos 1 e 2 da metodologia já foram determinados, nomeadamente a recolha dos dados e a sua respetiva análise preliminar.

No passo 3 é selecionada a distribuição que permite modelar os dados recolhidos.

Para escolher uma distribuição que se afigure como a mais indicada para representar os dados, é necessário ter em consideração a caracterização da informação.

O critério utilizado na escolha da distribuição foi, pelo conhecimento do comportamento do componente em análise e de outros componentes mecânicos, com comportamento semelhante e que foram alvo de estudo em outros trabalhos (Abernethy, 2006), (Ojile, 2010), (Rinne, 2009), (Wang, 2004).

Assim a distribuição escolhida, que presumivelmente melhor se ajusta aos dados, recaiu pela distribuição de Weibull devido à sua grande flexibilidade, isto é, engloba funções com intensidade de falha constantes, crescentes e decrescentes, dependendo do valor do parâmetro de forma (Rinne, 2009), (Tobias, 2011), conforme referido no capítulo 3.

Uma vez identificada a distribuição que melhor representa os dados recolhidos das falhas, segue-se o passo 4, que corresponde à estimação dos respetivos parâmetros da distribuição.

5.3.1.1 - Estimação dos parâmetros da distribuição

Os métodos utilizados foram, o método dos mínimos quadrados conjuntamente com o método gráfico e o método de máxima verosimilhança. A apresentação teórica de cada um dos métodos foi feita no capítulo 4.

a) Método dos mínimos quadrados

A construção gráfica que se apresenta, além de servir para estimar os parâmetros da distribuição assumida tem, também, a particularidade de permitir avaliar a qualidade do ajuste entre os dados e a distribuição em questão. Se estes surgirem como uma reta, poder-se-á depreender que existe um bom ajustamento entre os dados e a distribuição assumida.

Atendendo à caracterização da informação relativa ao componente, verifica-se que a maioria das falhas ocorreram antes da conclusão da recolha dos dados, proporcionando observações completas, porém, no último registo, as bombas continuaram em funcionamento para além desse período, assim a informação obtida é censurada à direita.

Conforme indicado no capítulo 4, com este tipo de dados, um dos métodos que melhor se adequa para a obtenção de $F(t)$, é o método de Herd-Johnson e desenrola-se sequencialmente através dos seguintes passos:

- ordenar os tempos de falha, t_i , por ordem crescente

- determinar $\begin{cases} \hat{R}_{f,j} = \left(\frac{n+1-l_j}{n+2-l_j} \right) \hat{R}_{f,(j-1)} \\ \hat{F}_{f,j} = 1 - \hat{R}_{f,j} \end{cases}$, conforme referido na equação 4.9
- calcular $x_i = \ln(t_i)$
- calcular $y_i = \ln[-\ln(1 - \hat{F}_i)]$
- representar o gráfico (x_i, y_i)

Considerando uma amostra de tamanho n com r falhas ($0 < r < n$) e $(n - r)$ dados censurados.

A figura seguinte representa o gráfico de probabilidade de Weibull logaritmizado, dos dados de falha, em horas, para um empanque mecânico.

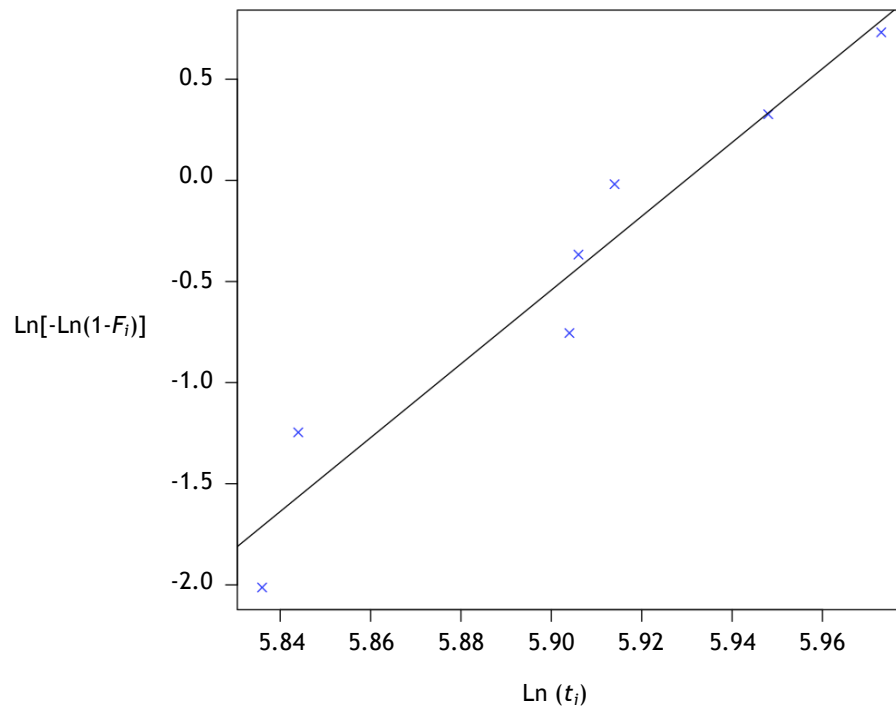


Figura 5.7 - Gráfico de probabilidade de Weibull dos dados de falha em horas para um empanque mecânico.

A figura 5.8 representa o gráfico de probabilidade de Weibull logaritmizado, dos dados de falha, em dias, para todos os empanques mecânicos.

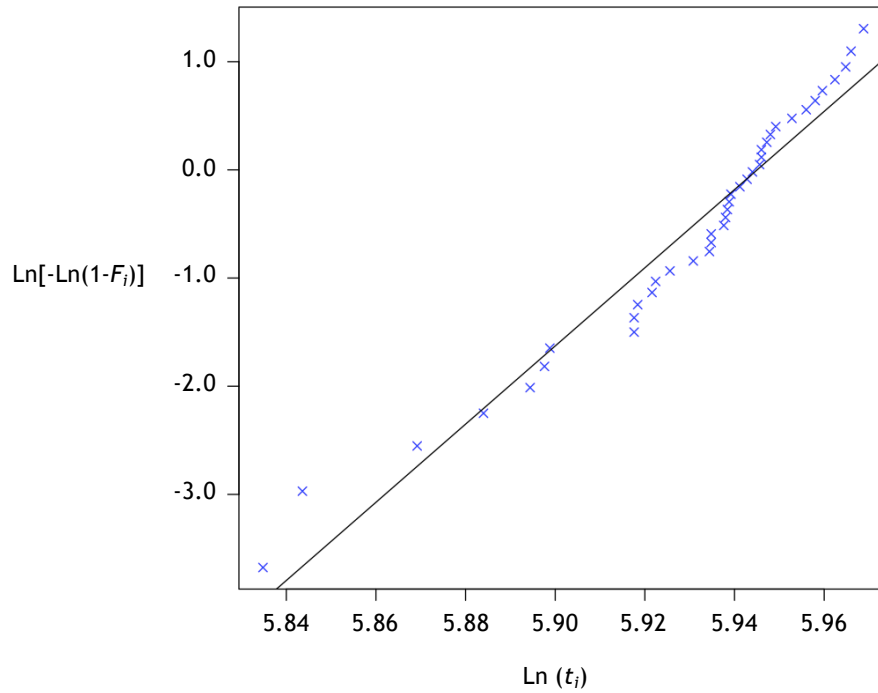


Figura 5.8 - Gráfico de probabilidade de Weibull dos dados de falha em dias para todos os empanques mecânicos.

Pela análise às figuras 5.7 e 5.8, verifica-se que existe um bom ajustamento entre os dados e a distribuição escolhida. Assim pode-se assumir que a distribuição de Weibull ajusta-se aos tempos de falha do empanque mecânico.

Através da equação 4.17 obtêm-se os valores estimados de $\hat{\beta}$ e de $\hat{\eta}$ pelo método dos mínimos quadrados,

$$\left\{ \begin{array}{l} \hat{\beta} = \frac{\sum_{i=1}^r [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^r (x_i - \bar{x})^2} = \frac{r \sum_{i=1}^r (x_i y_i) - \sum_{i=1}^r x_i \cdot \sum_{i=1}^r y_i}{r \sum_{i=1}^r x_i^2 - \left(\sum_{i=1}^r x_i \right)^2} \\ \hat{\eta} = \exp \left(-\frac{\bar{y} - \hat{\beta} \bar{x}}{\hat{\beta}} \right) = \exp \left(-\frac{\sum_{i=1}^r y_i - \hat{\beta} \sum_{i=1}^r x_i}{r \hat{\beta}} \right) \end{array} \right.$$

Na tabela 5.3 é apresentado o resultado do valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ para cada um dos empanques mecânicos.

Tabela 5.3 - Valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ (dias) para cada um dos empanques mecânicos pelo método dos mínimos quadrados.

| Empanque mecânico | $\hat{\beta}$ | $\hat{\eta}$ (dias) |
|-------------------|---------------|---------------------|
| 1 | 8,76 | 388,72 |
| 2 | 4,26 | 338,17 |
| 3 | 12,54 | 415,15 |
| 4 | 7,68 | 386,71 |
| 5 | 6,55 | 363,42 |

Todos os empanques mecânicos apresentam o parâmetro de forma, β maior do que 1, que significa que a função intensidade de falha é monótona crescente. Os valores do parâmetro de escala, η , são muito próximos entre os vários empanques, os quais variam entre 338,17 dias e 415,15 dias de funcionamento.

b) Método de máxima verosimilhança

O outro método utilizado na estimação dos parâmetros da distribuição de Weibull foi o método de máxima verosimilhança.

Para obter os valores estimados pelo método de máxima verosimilhança utilizou-se o algoritmo EM para dados censurados à direita, de forma diferente do habitual, tornando este procedimento original e distinto dos trabalhos publicados por outros autores (Balakrishnan, 2007), (Balakrishnan, 2012), conforme descrito no capítulo 4.2.2.2.2.1.

No processo de implementação do algoritmo foram utilizados os programas estatísticos R e Matlab, com respectivas livrarias e packages.

Na tabela seguinte, é apresentado o valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ para cada um dos empanques mecânicos pelo método de máxima verosimilhança com resolução através do algoritmo EM. São de igual modo apresentados os respetivos intervalos de confiança.

Como o número de dados da amostra é reduzido optou-se pela utilização do método bootstrap na determinação dos intervalos de confiança conforme recomendado no capítulo 4.3.3, e especificamente pelo método bootstrap-t conforme descrito no capítulo 4.3.3.2 com um coeficiente de confiança de 95%.

Todos os empanques mecânicos apresentam o parâmetro de forma, β maior do que 1, do mesmo modo que os valores obtidos pelo método dos mínimos quadrados. Os valores do parâmetro de escala, η , variam entre 345,31 dias e 403,68 dias de funcionamento.

Verifica-se que os valores obtidos do parâmetro de forma, β , pelo método de máxima verosimilhança são sempre maiores, do que, pelo método dos mínimos quadrados. Para o parâmetro de escala, η , os valores obtidos pelos dois métodos são próximos.

Tabela 5.4 - Valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ (dias) para cada um dos empanques mecânicos pelo método de máxima verosimilhança (EM) e respectivo intervalo de confiança pelo método bootstrap-t.

| Empanque mecânico | $\hat{\beta}$ | | | $\hat{\eta}$ (dias) | | |
|----------------------|---------------|-------|-------|---------------------|--------|--------|
| 1 | 6,12 | 9,21 | 14,09 | 365,51 | 392,96 | 425,53 |
| 2 | 2,01 | 5,04 | 10,12 | 301,49 | 345,31 | 392,67 |
| 3 | 10,85 | 13,92 | 17,19 | 385,03 | 403,68 | 423,21 |
| 4 | 4,93 | 8,17 | 12,98 | 352,85 | 381,44 | 416,05 |
| 5 | 3,96 | 7,56 | 12,51 | 335,62 | 368,37 | 406,99 |

O número de reamostragem para o método de bootstrap é igual a 1000. Na figura 5.9 é possível verificar a dispersão dos valores obtidos pelo método bootstrap para o empanque 1.

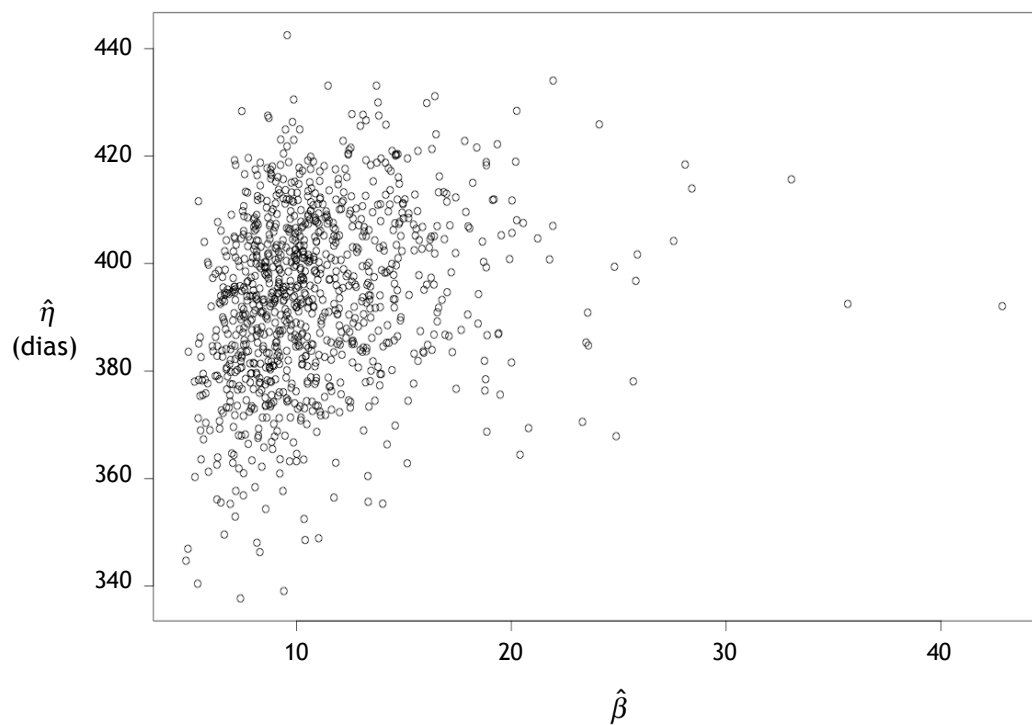


Figura 5.9 - Dispersão dos valores obtidos pelo método bootstrap para o empanque 1.

Com a informação obtida pelo método de bootstrap, pode-se representar o intervalo de confiança em torno da função densidade de probabilidade da distribuição Weibull estimada, conforme indicado na figura 5.10, para o empanque 1.

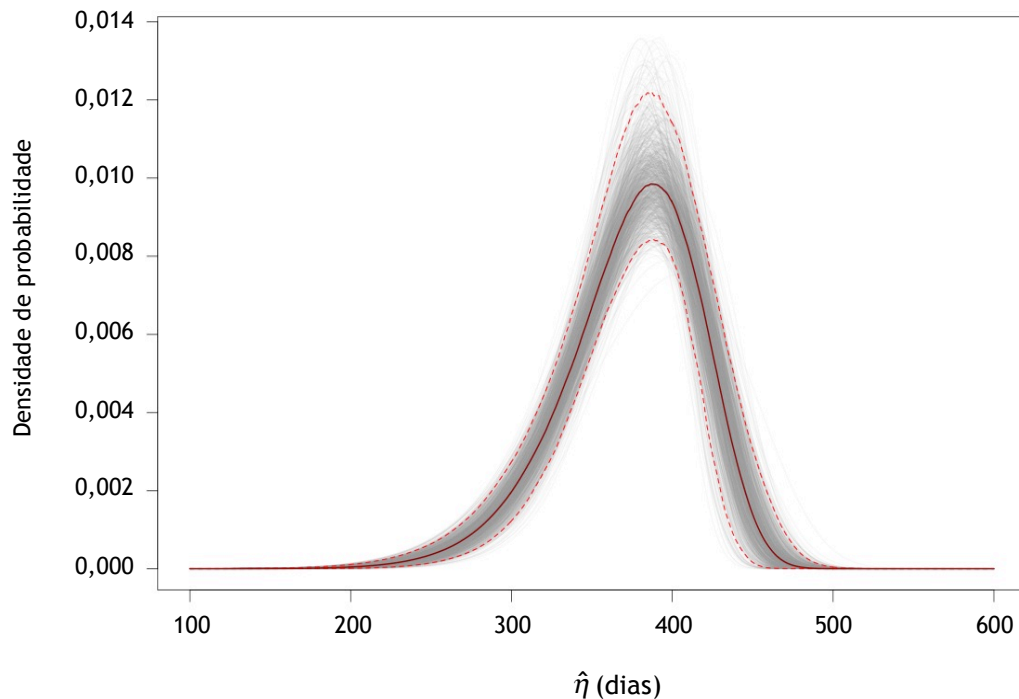


Figura 5.10 - Densidade de probabilidade pelo método bootstrap para o empanque 1.

Na figura 5.10 pode-se observar a evolução do intervalo de confiança ao longo da curva de densidade de probabilidade pelo método bootstrap, para o empanque 1. Note-se que no ponto de maior densidade, o limite superior do intervalo de confiança afasta-se mais do que o limite inferior. No entanto, o gráfico permite concluir que os dados não estão muito dispersos.

Existem vários testes que permitem completar a análise da qualidade do ajuste com o intuito de avaliar o ajustamento entre os dados considerados e a distribuição teórica. Os testes mais comuns são o teste do Qui-Quadrado e o teste de Kolmogorov-Smirnov (O'Connor, 2012), (Kececioglu, 2002).

O teste utilizado neste estudo foi o teste de Kolmogorov-Smirnov (K-S) porque é mais simples de aplicar e apresenta resultados mais fiáveis para amostras com quantidade reduzida de dados (O'Connor, 2012), (Kececioglu, 2002).

O teste K-S baseia-se na comparação da função de distribuição empírica com a função de distribuição teórica.

O método processa-se em 3 etapas.

1. Formulação das hipóteses

Na hipótese nula é especificado que a população segue uma determinada distribuição de probabilidade, em confronto com a hipótese alternativa, onde se especifica que a população não segue tal distribuição de probabilidade.

- H_0 : a população da qual provêm os dados segue a distribuição de Weibull;
- H_1 : a população da qual provêm os dados não segue a distribuição de Weibull.

2. Determinação do valor de D

Calcula-se o supremo da diferença em valor absoluto entre a função distribuição da amostra $F(t_i)$ e a função distribuição populacional $F(t)$.

Com o maior valor dessa diferença compara-se com o valor crítico.

O valor crítico obtém-se da tabela apresentada no anexo c, em função da dimensão da amostra, n , e do nível de significância, α . O valor obtido deve ser corrigido em função do valor do parâmetro de forma, β (O'Connor, 2012).

$$K = 0,70 \quad (\beta > 3.0)$$

$$K = 0,75 \quad (1,5 < \beta < 3)$$

$$K = 0,8 \quad (\beta < 1,5)$$

3. Comparação

Para um dado tamanho da amostra, n e um nível de significância, rejeita-se a hipótese nula quando o valor da estatística de teste, D , é maior ou igual que o valor crítico corrigido, isto é,

$$D \geq k \times \text{valor crítico} \Rightarrow \text{Rejeitar } H_0 \quad 5.1$$

O valor de prova de p-valor também permite avaliar a qualidade do ajuste.

O p-valor é a probabilidade de observar resultados tão extremos quanto aqueles que foram obtidos se a hipótese nula for verdadeira. A ideia é que se o p-valor for grande ele fornece evidência de que H_0 é verdadeira, enquanto que um p-valor pequeno indica que existe evidência nos dados contra H_0 . Se p-valor for superior a 0,05 não existe evidência contra H_0 (Pandis, 2013).

Para uma amostra $n = 7$ e com nível de significância $\alpha = 5\%$ referente ao empanque 1, vem,

$$D = 0,211 < 0,338 \quad (0,70 \times 0,483)$$

$$\text{p-valor} = 0,6148 > 0,05$$

Assim a hipótese nula não é rejeitada, isto é, não se rejeita a hipótese de que a população da qual provêm os dados segue a distribuição de Weibull.

Para os outros empanques os resultados formam semelhantes.

5.3.2 - Rolamentos

Conforme indicado anteriormente, numa bomba centrífuga, os rolamentos são o segundo componente responsável por mais falhas.

O rolamento é composto por dois anéis metálicos concêntricos nos quais se encontram as pistas, uma no anel externo e outra no anel interno, por um elemento espaçador ou gaiola e pelos corpos rolantes (rolos ou esferas) (Bloch, 1998), (Forsthoffer, 2005).

Os rolamentos têm a função de suporte do veio da bomba centrífuga. Suportam as cargas estáticas e dinâmicas existentes nas partes móveis do equipamento. Os rolamentos operam sob rotação e mantêm as posições relativas entre as partes rotativas e estacionárias.

O tempo de vida de um rolamento é definido como o número de rotações que este pode efetuar antes de aparecerem os primeiros sinais de desgaste, na forma de fissuras (cracks). O desgaste por fadiga é o resultado de tensões cíclicas que aparecem imediatamente abaixo da superfície carregada. Após algum tempo, estas tensões provocam microfissuras que, gradualmente, evoluem até atingirem a superfície. Com a passagem dos corpos rolantes desprendem-se pequenos fragmentos, fenómeno conhecido como *flaking* ou *spalling*. A área inicial danificada é normalmente muito pequena. No entanto, maiores tensões, em combinação com os fragmentos libertados no lubrificante, provocam a evolução da área danificada. Este processo ocorre num período de tempo relativamente longo e deteta-se a sua evolução pelo aumento do ruído e da vibração do rolamento (Bachus, 2003), (Bloch, 2010).

Os avanços da tecnologia na metalurgia e na lubrificação criaram condições para utilizar um filme de óleo cada vez mais fino entre os corpos rolantes e as pistas dos rolamentos, para minimizar a fricção, maximizar o desempenho e aumentar drasticamente a vida dos rolamentos (Bloch, 1999).

Infelizmente, acontece por vezes que o tempo de vida de um rolamento não coincide com o previsto, havendo muitas razões para tal, como por exemplo, carga inadequada, deficiente lubrificação, vibrações, problemas de alinhamento, folgas demasiado apertadas, etc..

Na referência bibliográfica (Bloch, 2010) é apresentado um estudo onde se indica as principais causas de falha nos rolamentos de uma bomba centrífuga, conforme indicado na figura seguinte.

De acordo com o estudo apresentado menos de 10% dos rolamentos funcionam o tempo suficiente para sucumbir por fadiga.

Verifica-se que as falhas relacionadas com a lubrificação são responsáveis por 54% de todas as falhas nos rolamentos.

A lubrificação do rolamento tem as seguintes funções:

- Formar um filme fluido entre os corpos rolantes e as pistas, de modo a minimizar o contato metal-metal, o atrito e o desgaste.
- Conter aditivos químicos para minimizar a corrosão na superfície de contato.
- Reduzir o calor gerado pelo atrito interno do rolamento através de trocas térmicas.

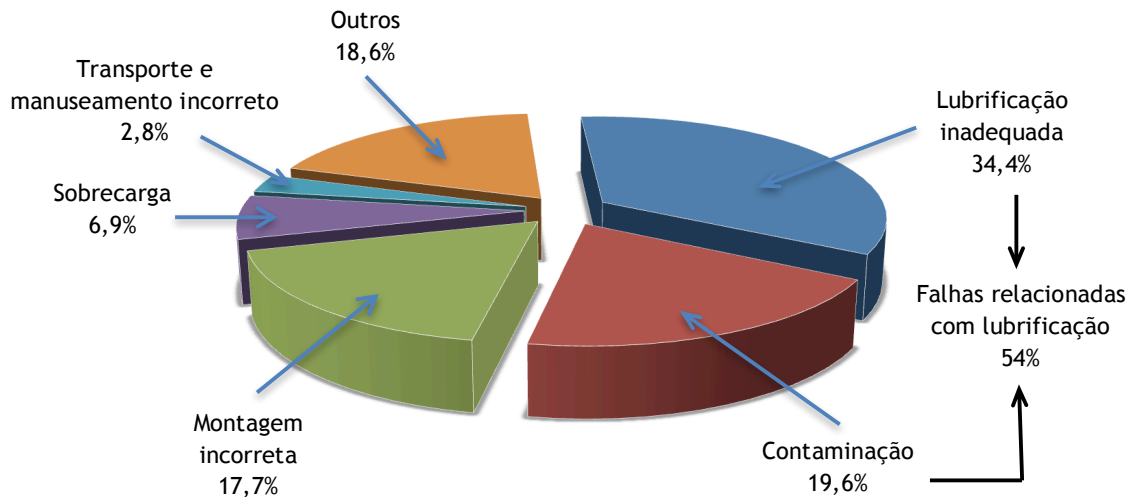


Figura 5.11 - Principais causas de falha no rolamento de uma bomba centrífuga (Bloch, 2010).

Conforme referido anteriormente, uma das formas de contaminação do lubrificante é através da libertação de fragmentos dos corpos rolantes e das pistas dos rolamentos devido ao processo de desgaste a que está submetido.

Assim a falta de lubrificação, a escolha incorreta do lubrificante ou a sua contaminação reduz significativamente o tempo de vida do rolamento.

Outras causas levam ao surgimento de falhas nos rolamentos, como por exemplo (Bachus, 2003), (Karassik, 2001):

- Cavitação: Um dos motivos da vibração é a cavitação. A cavitação é um fenómeno muito conhecido e ocorre quando a pressão na sucção da bomba se torna tão baixa, a ponto de permitir que o fluido bombeado se vaporize na entrada do impulsor. À medida que o líquido vaporizado se encaminha para regiões de pressão mais elevadas ainda dentro do próprio impulsor, ocorre nova mudança de fase, no sentido inverso, o que ocasiona elevadas vibrações na bomba devido ao colapso das bolhas. Além disso, é muito comum acontecer o arrancamento de material do impulsor e da carcaça, causado pelo referido colapso das bolhas.
- Desequilíbrio: Este problema ocorre quando o eixo principal de inércia do rotor não coincide com a sua direção de rotação. Todas as máquinas apresentam desequilíbrio, sendo a questão principal identificar qual o nível tolerado pelo sistema. O principal efeito do desequilíbrio é aumentar significativamente a vibração global do equipamento. Se o desequilíbrio for muito acentuado, é possível observar, também, aquecimento dos rolamentos, o que significa que uma potencial falha está muito próxima.

- Veio empenado: O principal sintoma apresentado por uma bomba com eixo empenado é o aumento de vibração global da máquina, que é bastante alta na direção axial do eixo. Quanto mais ao centro do eixo, mais próximo do comportamento de um desequilíbrio.
- Desalinhamento: O desalinhamento apresenta vibrações elevadas nas direções axial e radial, podendo também apresentar aquecimento dos rolamentos. Entretanto, é comum que as vibrações nas direções axial e radial se apresentem em níveis muito distintos.
- Ressonância: Este fenômeno ocorre quando uma certa força oscilatória excita a frequência natural do rotor. O seu efeito é um aumento acentuado da vibração global, com espectro concentrado em 1x a rotação da máquina. É importante destacar também que o desgaste dos rolamentos geralmente contribui para a entrada em ressonância de uma bomba, dado que a frequência da vibração correspondente a essa causa tende a se aproximar da frequência natural da rotação da bomba.
- Corrosão: Os limites de resistência à fadiga são sensivelmente alterados em meios corrosivos, podendo os danos por fadiga ocorrer sob regimes de carga ainda menores. O processo corrosivo ocorre independente do regime de operação ao qual o equipamento está submetido. As bombas centrífugas que trabalham sob condições ambientais propícias à corrosão devem ser especialmente preparadas, seja pela aplicação de ligas metálicas resistentes ou pela aplicação de revestimentos de proteção

Na tabela 5.5 são indicados os modos mais comuns de falha, respectivos mecanismos e causas no rolamento de uma bomba centrífuga.

5.3.2.1 - Estimação dos parâmetros da distribuição

As considerações assumidas no caso anterior mantêm-se neste estudo. As falhas são consideradas completas exceto a última falha de cada rolamento que é considerada como falha censurada à direita, porque as bombas em estudo continuaram em funcionamento quando terminou o período de recolha de dados.

Também para este caso, a distribuição escolhida para representar os tempos de falhas foi a distribuição de Weibull, pelos motivos referidos no caso do empaque mecânico.

Com estas considerações é possível caracterizar o componente em análise e proceder à determinação dos parâmetros da distribuição estatística que melhor se ajusta aos tempos de falha dos rolamentos.

Tabela 5.5 - Modos, causas e respectivos mecanismos de falha nos rolamentos (Bloch, 1999).

| Modo de falha | Mecanismo de falha | Causa da falha |
|---------------|--------------------------------|-------------------------|
| Desgaste | - Desgaste da gaiola | Montagem incorreta |
| | | Contaminação |
| | | Falta de lubrificação |
| | | Inadequada lubrificação |
| | - Desgaste abrasivo/ desgaste | Montagem incorreta |
| | | Erro de projeto |
| | | Vibração |
| | | Contaminação |
| | | Falta de lubrificação |
| | | Inadequada lubrificação |
| | Sobreaquecimento -> scuffing | Sobrecarga |
| | | Falta de lubrificação |
| | | Lubrificação excessiva |
| | | Lubrificação inadequada |
| Deformação | - Deformação da gaiola | Montagem incorreta |
| | | Desalinhamento |
| | - Brinelling/ false brinelling | Contaminação |
| | | Montagem incorreta |
| | | Vibração |
| | | Lubrificação inadequada |
| | - Deformação da pista | Montagem incorreta |
| | | Sobrecarga |
| Fratura | - Apreensão | Erro de projeto |
| | | Lubrificação excessiva |
| | | Lubrificação inadequada |
| | | Falta de lubrificação |
| | - Fratura da gaiola | Desalinhamento |
| | | Vibração |
| | | Falta de lubrificação |
| | - Spalling (fragmentação) | Montagem incorreta |
| | | Vibração |
| | | Fadiga |
| | | Sobrecarga |
| | - Crack/ heat crack | Montagem incorreta |
| | | Erro de projeto |
| | | Sobrecarga |

a) Método dos mínimos quadrados

Conforme indicado no caso anterior, com este tipo de dados, um dos métodos que melhor se adequa para a obtenção de $F(t)$, é o método de Herd-Johnson e desenrola-se sequencialmente através dos passos indicados.

A figura 5.12 representa o gráfico de probabilidade de Weibull logaritmizado, dos dados de falha, em horas, para os rolamentos de uma bomba centrífuga.

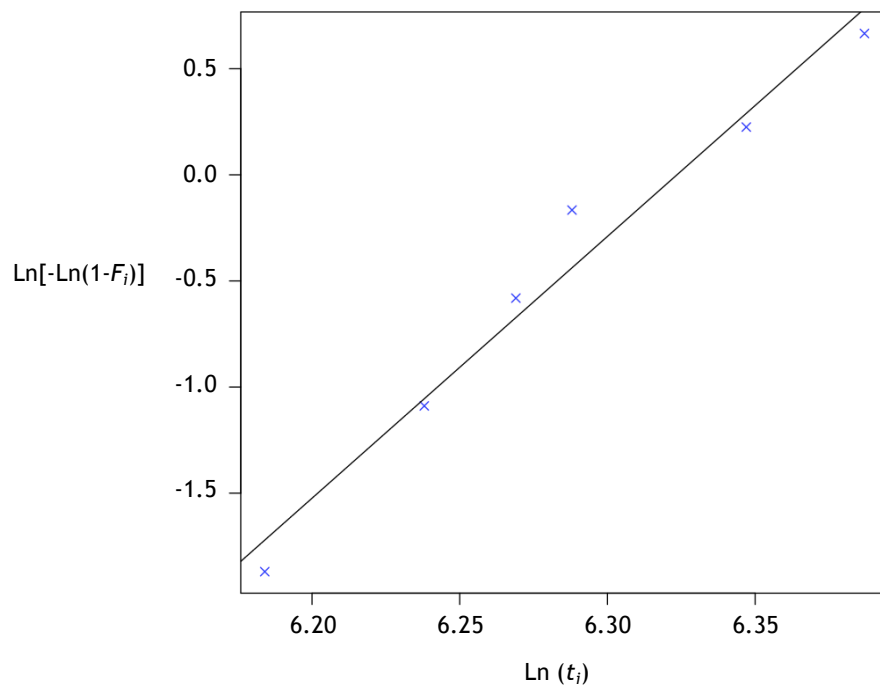


Figura 5.12 - Gráfico de probabilidade de Weibull dos dados de falha em horas para um rolamento.

A figura 5.13 representa o gráfico de probabilidade de Weibull logaritmizado dos dados de falha, em dias, para todos os rolamentos.

Pela análise às figuras 5.12 e 5.13 verifica-se que existe uma boa qualidade de ajuste entre os dados e a distribuição escolhida. Assim, pode-se assumir que a distribuição de Weibull ajusta-se aos tempos de falha dos rolamentos.

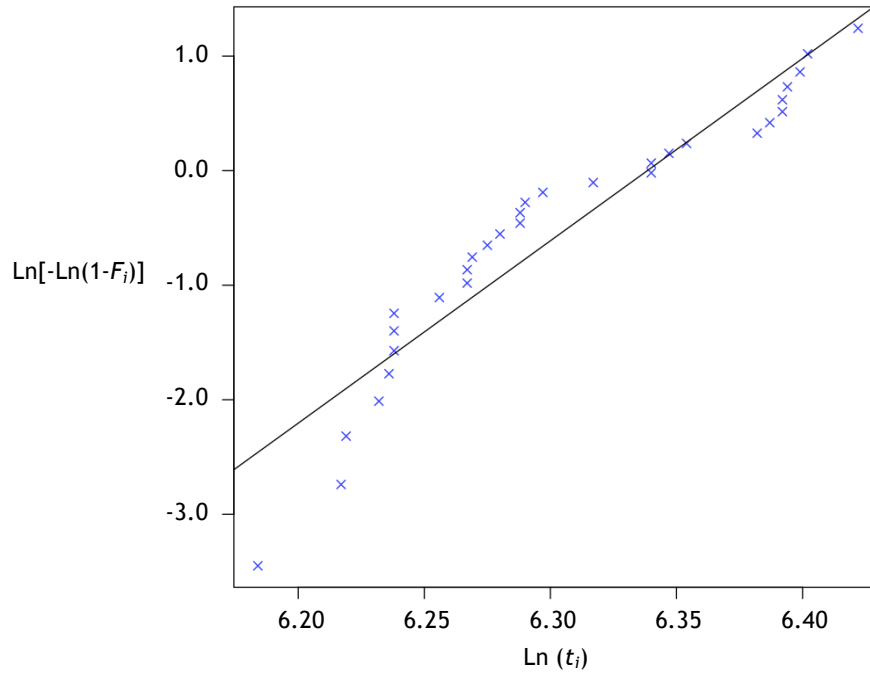


Figura 5.13 - Gráfico de probabilidade de Weibull dos dados de falha em dias para todos os rolamentos.

Na tabela 5.6 é apresentado o resultado do valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ dos rolamentos para cada uma das bombas centrífugas.

Tabela 5.6 - Valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ (dias) dos rolamentos para cada bomba centrífuga pelo método dos mínimos quadrados.

| Rolamentos | $\hat{\beta}$ | $\hat{\eta}$ (dias) |
|------------|---------------|---------------------|
| 1 | 7,14 | 567,59 |
| 2 | 6,91 | 519,41 |
| 3 | 9,02 | 583,54 |
| 4 | 5,34 | 537,78 |
| 5 | 3,03 | 648,56 |

Todos os rolamentos apresentam o parâmetro de forma, β , maior do que 1 que significa que a função intensidade de falha é monótona crescente. Os valores do parâmetro de escala, η , são muito próximos entre os vários rolamentos, os quais variam entre 519,41,04 dias e 648,56 dias de funcionamento.

b) Método de máxima verossimilhança

O outro método utilizado na estimação dos parâmetros da distribuição de Weibull foi o método de máxima verossimilhança.

No processo de implementação do algoritmo foram utilizados os programas estatísticos R e Matlab, com respectivas livrarias e packages.

Na tabela 5.7 é apresentado o valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ para cada um dos rolamentos pelo método de máxima verossimilhança com resolução através do algoritmo EM. São de igual modo apresentados os respectivos intervalos de confiança pelo método bootstrap-t para um nível de confiança de 95%.

Tabela 5.7 - Valor esperado de $\hat{\beta}$ e de $\hat{\eta}$ (dias) dos rolamentos para cada bomba centrífuga pelo método de máxima verossimilhança (EM) e respectivo intervalo de confiança pelo método bootstrap-t.

| Rolamentos | $\hat{\beta}$ | | | $\hat{\eta}$ (dias) | | |
|------------|---------------|------|-------|---------------------|--------|--------|
| 1 | 5,67 | 8,34 | 14,34 | 544,75 | 584,81 | 626,10 |
| 2 | 4,15 | 7,22 | 12,75 | 476,28 | 509,95 | 551,36 |
| 3 | 7,19 | 9,98 | 15,11 | 579,78 | 601,29 | 631,04 |
| 4 | 3,91 | 6,97 | 11,86 | 493,96 | 529,86 | 572,89 |
| 5 | 2,77 | 5,74 | 10,78 | 611,07 | 656,96 | 701,03 |

Todos os rolamentos apresentam o parâmetro de forma, β maior do que 1, do mesmo modo que os valores obtidos pelo método dos mínimos quadrados.

Os valores do parâmetro de escala, η , variam entre 509,95 dias e 656,96 dias de funcionamento.

Também neste caso, verifica-se que os valores obtidos do parâmetro de forma, β , pelo método de máxima verossimilhança são sempre maiores do que pelo método dos mínimos quadrados. Para o parâmetro de escala, η , os valores obtidos pelos dois métodos são próximos.

O número de reamostragem para o método de bootstrap é igual a 1000. Na figura 5.14 é possível verificar a dispersão dos valores obtidos pelo método bootstrap para o rolamento 1.

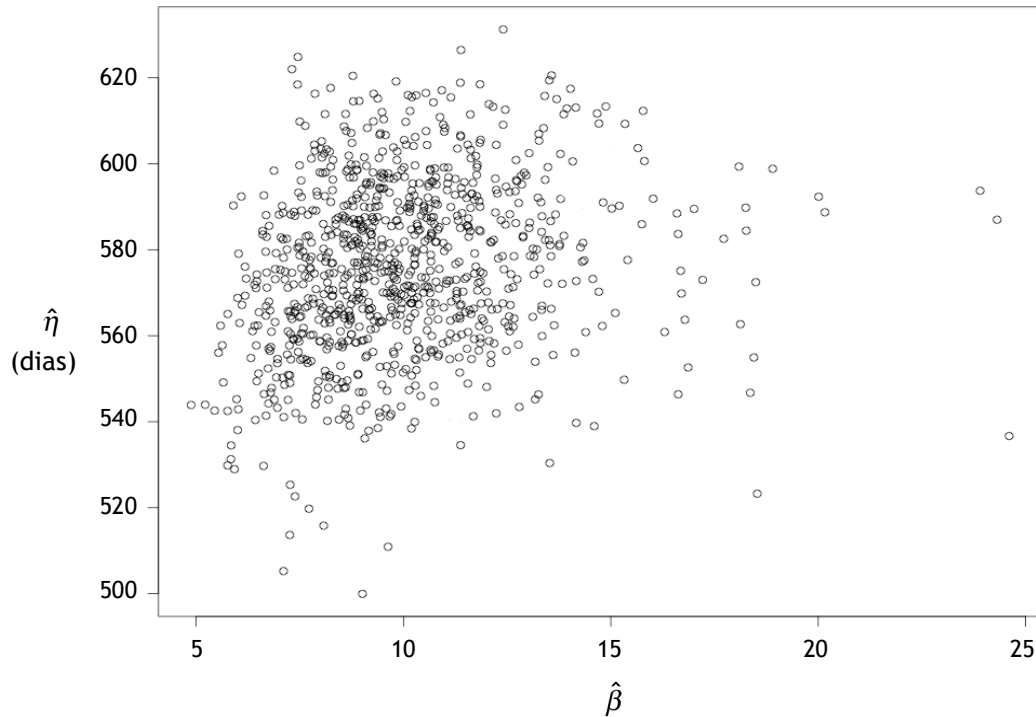


Figura 5.14 - Dispersão dos valores obtidos pelo método bootstrap para o rolamento 1.

Relativamente ao teste do ajuste dos dados utilizou-se o teste de Kolmogorov-Smirnov conforme indicado no caso do empanque mecânico.

Formulação das hipóteses:

- H_0 : a população da qual provêm os dados segue a distribuição de Weibull;
- H_1 : a população da qual provêm os dados não segue a distribuição de Weibull.

Para uma amostra $n = 6$ e com nível de significância $\alpha = 5\%$ referente ao rolamento 1, vem,

$$D = 0,218 < 0,363 (0,70 \times 0,519)$$

$$p\text{-valor} = 0,5294 > 0,05$$

Assim, a hipótese nula não é rejeitada, isto é, não se rejeita a hipótese de que a população da qual provêm os dados segue a distribuição de Weibull.

Para os outros rolamentos os resultados foram semelhantes.

5.4 - Conclusões

Neste capítulo conclui-se que os dois componentes com o maior registo de falhas numa bomba centrífuga são o empanque mecânico e os rolamentos. É também evidenciado que o registo histórico de falhas de sistemas industriais apresenta muitas vezes um número reduzido de dados e dados censurados, como por exemplo, dados censurados à direita.

Assim, a partir da apreciação destes resultados, justifica-se a importância de um estudo mais pormenorizado que consiga encontrar soluções para responder a essas dificuldades reais.

Da análise feita deste capítulo, verifica-se a aplicabilidade do algoritmo EM na estimação dos parâmetros da distribuição de Weibull e do método bootstrap para a determinação do intervalo de confiança na presença de um número reduzido de dados.

Da análise comparativa dos resultados obtidos pelo método dos mínimos quadrados e pelo método de máxima verosimilhança, através do algoritmo EM, constata-se que o parâmetro de forma, β , é maior pelo método de máxima verosimilhança do que pelo método dos mínimos quadrados e em ambos é superior a 1.

Relativamente ao parâmetro de escala, η , os valores obtidos pelos dois métodos são próximos.

Quando é necessário estimar os parâmetros de um modo simples e com boa aproximação, recomenda-se a utilização do método dos mínimos quadrados conjuntamente com o método gráfico, e assim também permite a obtenção de uma confirmação visual rápida do ajustamento de um determinado conjunto de dados.

Quando é necessária uma maior precisão do valor estimado, recomenda-se a utilização do método de máxima verosimilhança, conforme indicado no capítulo 4.2.2.2.1.

Se as equações de máxima verosimilhança não apresentarem solução analítica, como é habitual ocorrer na presença de dados censurados, verifica-se que é válida a utilização do algoritmo Expectation-Maximization (EM).

O conhecimento do comportamento dos componentes permite reduzir significativamente os custos associados à manutenção pela otimização do número de inspeções, da diminuição do número de componentes substituídos e de stock, melhor preparação dos trabalhos de manutenção e diminuição dos tempos de reparação, por exemplo.

Capítulo 6

Conclusões

O principal objetivo desta tese consiste na estimação dos parâmetros da distribuição fiabilística que melhor se ajusta aos dados provenientes de um registo histórico, de modo a conhecer o comportamento do equipamento ou do componente, conforme referido no capítulo 2.

Foi evidenciado a importância da aplicação de métodos fiabilísticos de modo a conhecer o comportamento do equipamento ou do componente analisado como ferramenta de apoio à gestão da manutenção. Apresentando-se as vantagens de conjugar este procedimento nos diferentes tipos de manutenção.

Para tal, foi proposto uma metodologia composta em cinco etapas, que permite realizar o processo de estimação paramétrica da distribuição.

Pela comparação através da simulação de Monte Carlo, cujos resultados foram apresentados no capítulo 4, verificou-se que o método de máxima verosimilhança apresenta melhores resultados para dados completos e censurados, do que o método dos mínimos quadrados. Contudo, quando a precisão dos valores obtidos não é um elemento fundamental, o método dos mínimos quadrados, conjuntamente com a representação gráfica, apresenta-se como uma boa alternativa, devido à sua implementação simples, com relativa precisão dos valores obtidos e permite uma visualização gráfica do comportamento dos dados.

Relativamente à estimação de $F(t)$, que é necessário para o método gráfico e para o método dos mínimos quadrados, verificou-se através da simulação de Monte Carlo, que o método de Bernard apresenta melhores resultados na presença de dados completos e o método de Herd-Johnson apresenta melhores resultados na presença de dados censurados.

Foi referido a importância do histórico de falhas, apresentando-se como a melhor fonte de informação sobre o sistema em análise, contudo, os dados provenientes do registo histórico em contexto operacional apresentam a informação com características específicas que dificultam a sua análise, nomeadamente, dados incompletos e reduzido número de

dados. Neste contexto, as equações de máxima verosimilhança, podem nestas circunstâncias não apresentar solução analítica, sendo necessário recorrer a métodos numéricos.

Nesta tese, foi possível verificar a aplicação do algoritmo EM como método numérico para determinar as soluções da equação que deriva do método da máxima verosimilhança, em particular, no caso da presença de dados censurados à direita.

Na presente tese utilizou-se o procedimento base do algoritmo EM, contudo, foi introduzida uma nova abordagem na sua aplicação, no caso específico de dados censurados à direita provenientes de um registo histórico, tornando este procedimento original e distinto dos trabalhos publicados por outros autores, conforme descrito no capítulo 4.

O algoritmo EM apresenta várias propriedades e vantagens, que o destacam relativamente a outros algoritmos iterativos, nomeadamente, o método de Newton-Raphson:

- O algoritmo EM converge sob condições bastante amplas, ou seja, a partir de um dado arbitrário, $\theta^{(0)}$, o algoritmo geralmente encontra um máximo local, com exceção de uma má escolha da solução inicial $\theta^{(0)}$ ou na formulação errada da função de verosimilhança.
- O trabalho analítico necessário é mais simples do que com outros métodos, dado que só é necessário maximizar o valor esperado condicional da log-verosimilhança para dados completos.
- O algoritmo EM é relativamente fácil de programar e de ser implementado.
- Durante as iterações é possível controlar a convergência e os erros de programação

Contudo também apresenta algumas desvantagens:

- O algoritmo EM pode convergir lentamente, mesmo em alguns problemas aparentemente simples, e nos problemas em que há muita informação incompleta.
- O algoritmo EM não tem um processo integrado para produzir uma estimativa da matriz de covariância dos parâmetros estimados. Contudo, esta desvantagem pode ser contornada pela utilização de metodologia adequada.
- O algoritmo EM, como o método de Newton-Raphson, não garante a convergência para o máximo global quando existem vários máximos locais. A estimativa obtida depende da solução inicial.

O processo de estimação foi aplicado em cinco bombas centrífugas de uma indústria petroquímica. Verificou-se que os dois componentes com o maior registo de falhas foram o empanque mecânico e os rolamentos. Conjuntamente, correspondem aproximadamente a 80% do número total de falhas registadas. Este resultado corrobora com outro estudo apresentado.

Da análise feita deste trabalho, com dados obtidos em contexto industrial, verificou-se que é possível utilizar a distribuição de Weibull na modelação do comportamento dos referidos componentes. Conforme referido no capítulo 3, a distribuição de Weibull apresenta propriedades que permitem caracterizar bem este tipo de sistema.

Os objetivos da aplicação da metodologia, neste sistema em análise, foram alcançados, designadamente, a modelação e desenvolvimento do algoritmo EM na estimação dos parâmetros da distribuição de Weibull e a utilização do método bootstrap para a determinação do intervalo de confiança na presença de um número reduzido de dados.

6.1 - Trabalhos futuros

Concluída a apresentação do trabalho desenvolvido ao longo do doutoramento, no âmbito da estimação dos parâmetros da distribuição de Weibull, importa, ainda, referir algumas sugestões, para desenvolvimento futuro, relacionadas com o assunto tratado.

No que respeita a sugestões de trabalhos futuros, importa referir o interesse da aplicabilidade do método e designadamente, do algoritmo EM, em outros sistemas e com outros modos de falhas.

Será interessante estudar a utilização do algoritmo EM com outros tipos de dados censurados e truncados, bem como, da sua presença em simultâneo.

Relativamente à modelização e simulação, de modo a validar o desempenho, será importante comparar com outras metodologias, nomeadamente, métodos Bayesianos.

Por ultimo, apresenta-se também, como uma sugestão de grande interesse para trabalho futuro, a aplicação do algoritmo EM na resolução numérica da função de máxima verosimilhança em outras distribuições estatísticas.

Referências

Abernethy, R. B. (2006), *The New Weibull Handbook*, 5th Ed, Robert B. Abernethy, ISBN 0-9653062-3-2, Florida, USA.

Ahmadi, J., Doostparast, M., Parsian, A. (2012), *Estimation with left-truncated and right censored data: a comparison study*, Statistics and Probability Letters, Vol. 82, pp. 1391-1400.

Ahmed, A. O. M. (2014), *Comparison of the bayesian methods on interval-censored data for Weibull distribution*, Open Journal of Statistics, Vol. 4, pp. 570-577.

Ahmed, S. A. (2013), *Comparative study of four methods for estimating Weibull parameters*, International Journal of Physical Sciences, Vol. 8, pp. 186-192.

Akram, M., Hayat, A. (2014), *Comparison of estimators of the Weibull distribution*, Journal of Theoretical and Applied Statistics, Vol. 8 (1), pp. 238-259.

Arasan, J., Adam, M. B. (2014), *Double Bootstrap confidence interval estimates with censored and truncated data*, Journal of Modern Applied Statistical Methods, Vol. 13 (2), pp. 399-419.

Ascher, H., Feingold H. (1984), *Repairable Systems Reliability - Modeling, Inference, Misconceptions and their causes*. Marcel Dekker, Inc., ISBN 0-8247-7276-8, New York, USA.

Assis, R. (2009), *Calendário de inspeções em manutenção preventiva condicionada com base na fiabilidade*, 3º Encontro Nacional de Riscos, Segurança e Fiabilidade, IST, Lisboa, Portugal.

Bachus, L., Custodio, A. (2003), *Know and understand centrifugal pumps*, Elsevier, ISBN 1856174093, Oxford, UK.

Balakrishnan, N., Kundu, D. (2013), *Hybrid censoring: models, inferential results and applications*, Computational Statistics and Data Analysis, Vol. 57, pp. 166-209.

Balakrishnan, N., Kundu, D., Ng, H. K. T. (2007), *Point and interval estimation for a simple step-stress model with Type-II censoring*, Journal of Quality Technology, Vol. 39, pp. 35-47.

Balakrishnan, N., Mitra, D. (2011), *Likelihood inference for lognormal data with left truncated and right censoring with an illustration*, Journal of Statistical Planning and Inference, Vol. 141, pp. 3536-3553.

Balakrishnan, N., Mitra, D. (2012), *Left truncated and right censored Weibull data and likelihood inference with an illustration*, Computational Statistics and Data Analysis, Vol. 56, pp. 4011-4025.

Birolini, A. (2014), *Reliability engineering theory and practice*, 7th Ed., Springer, ISBN 978-3-642-39534-5, London, UK.

Bloch, Heinz P. (1998), *Practical machinery management for process plant, improving machinery reliability, volume 1*, 3rd Ed., Elsevier, ISBN 0-88415-661-3, New York, USA.

Bloch, Heinz P., Budris, Allan R. (2010), *Pump user's handbook life extension*, 3rd Ed., CRC Press, ISBN 978-1-4398-3740-5, New York, USA.

Bloch, Heinz P., Geitner, Fred K. (1999), *Practical machinery management for process plant, machinery failure analysis and troubleshooting, volume 2*, 3rd Ed., Elsevier, ISBN 0-88415-662-1, New York, USA.

Bohning, D. (1999), *Computer-Assisted Analysis of Mixtures and Applications: Meta-analysis, Disease Mapping and Others*, Chapman & Hall, ISBN 978-1-5848-8179-7, New York, USA.

Cabral, J. S. (1998), *Organização e Gestão da Manutenção*, Lidel, Lisboa, Portugal.

Carpenter, J., Bithell, J. (2000), *Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians*, Statistics in Medicine, Vol. 19, pp. 1141-1164.

Chambers, R. L., Steel, D. G., Wang, S., Welsh, A. H. (2012), *Maximum likelihood estimation for sample surveys*, CRC Press, ISBN 978-1-4200-1135-7, Florida, USA.

Cheng, C., Chen, J., Li, Z. (2013), *A new algorithm for maximum likelihood estimation with progressive type-I interval censored data*, Communications in Statistics - Simulation and Computation, Vol. 39, pp. 750-766.

Cohen, A. C. (1991), *Truncated and censored samples*, Marcel Dekker, ISBN 0-8247-8447-2, New York, USA.

Davison, A. C., Hinkley, D. V. (1997), *Bootstrap methods and their application*, Cambridge University Press, ISBN 978-0-521-57391-7, Cambridge, UK.

Dekker R., Scarf P. A. (1998), *On the impact of optimisation models in maintenance decision making: the state of the art*, Reliability Engineering and System Safety, vol. 60, pp. 111-119.

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977), *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society, vol. 39 (1), pp. 1-38.

Di Febo, M., Paganini, P. (2015), *Centrifugal pump technology in oil and gas refinement*, World Pumps, Vol. 2015 (11), pp. 36-40.

Dodson, B. (2006), *The Weibull Analysis Handbook*, 2nd Ed. American Society for Quality, Quality Press, ISBN 978-0-87389-667-2, Milwaukee, USA.

Dubi, A. (2000), *Monte Carlo Applications in Systems Engineering*, John Wiley & Sons, ISBN 0-471-98172-9, Chichester, UK.

Ebeling, C., Lewis (1997), *An Introduction to Reliability and Maintainability Engineering*, McGraw-Hill, New York, USA.

Edwards, A. W. F. (1974), *The history of likelihood*, International Statistical Institute, Vol. 42 (1), pp. 9-15.

Efron, B. (1979), *Bootstrap methods: Another look at the Jackknife*, The Annals of Statistics Vol. 7 (1), pp. 1-26.

Efron, B., Tibshirani, R. J. (1986), *Bootstrap methods for standart errors, confidence intervals and other measures of statistical accuracy*, Statistical Science, Vol. 1 (1), pp. 54-75.

Efron, B., Tibshirani, R. J. (1993), *An introduction to the Bootstrap*, Chapman & Hall, ISBN 0-412-04231-2, Londres, UK.

Elmahdy, E. E., Aboutahoun, A. W. (2013), *A new approach for parameter estimation of finite Weibull mixture distributions for reliability modeling*, Applied Mathematical Modelling, Vol. 37, pp. 1800-1810.

Fang, L. Y., Arasan, J., Midi, H., Bakar, M. R. A. (2015), *Jackknife and Bootstrap inferential procedures for censored survival data*, AIP Conference Proceedings, Vol. 1682, pp. 1-6.

Faria, S., Soromenho, G. (2010), *Fitting mixtures of linear regressions*, Journal of Statistical Computation and Simulation, vol. 80 (2), pp. 201-225.

Ferreira, L.A. (1998), *Uma Introdução à Manutenção*, Publindústria, ISBN 972-95794-4-X, Porto, Portugal.

Ferreira, L. A. (2000), *Engenharia de manutenção*, Actas do 2º Encontro Nacional do Colégio de Engenharia Mecânica da Ordem dos Engenheiros de Portugal, Coimbra, Portugal.

Forsthoffer, William E. (2005), *Forsthoffer's Rotating Equipment Handbooks: Pumps*, Volume 2, Elsevier, ISBN 1-85617-468-9, Oxford, UK.

Fothergill, J. C. (1990), *Estimating the cumulative probability of failure data points to be plotted on Weibull and other probability paper*, IEEE Transactions on Electrical Insulation, Vol. 25 (3), pp. 489-492.

Gijbels, I. (2010), *Censored data*, John Wiley & Sons, Wires Computational Statistics, Vol. 2, pp. 178 - 188.

Girdhar, P., Moniz, O. (2005), *Practical centrifugal pumps - design, operation and maintenance*, Elsevier, ISBN 0-7506-6273-5, Oxford, UK.

Guimarães, R. C., Cabral, J. A. S. (1999), *Estatística*, McGraw-Hill, ISBN 972-8298-45-5, Amadora, Portugal.

Guure, C. B., Ibrahim, N. A. (2012), *Bayesian analysis of the survival function and failure rate of Weibull distribution with censored data*, Hindawi Publishing Corporation, Mathematical Problems in Engineering, Vol. 2012, pp. 1-18.

Guure, C. B., Ibrahim, N. A. (2013), *Methods for estimating the 2-parameter Weibull distribution with type-I censored data*, Journal of Applied Sciences, Engineering and Technology Vol. 5 (3), pp 689-694.

Hall, M. J., Van den Boogaard, H. F. P., Fernando, R. C., Mynett, A. E. (2004), *The construction of confidence intervals for frequency analysis using resampling techniques*, Hydrology and Earth System Sciences, Vol. 8 (2), pp. 235-246.

Held, L., Bové, D. (2014), *Applied Statistical Inference, Likelihood and Bayes*, Springer, ISBN 978-3-642-37886-7, Berlin, Germany.

Hossain, A. M., Zimmer, W. J. (2003), *Comparisons of estimation methods for Weibull parameters: complete and censored samples*, Journal of Statistical Computation and Simulation, Vol. 73 (2), pp. 145-153.

Ibrahim, N. A., Kudus, A. (2007), *Bootstrap confidence interval for the median failure time of three-parameter Weibull distribution*, Proceedings of the World Congress on Engineering, Vol. II, pp. 836-839.

Jardine, Andrew K. S., Tsang, Albert H. C. (2013), *Maintenance, replacement and reliability: theory and applications*, 2st Ed., CRC Press, ISBN 978-1-4665-5486-3, New York, USA.

Kalbfleisch, J. D., Prentice, R. L. (2002), *The statistical analysis of failure time data*, John Wiley & Sons, Ltd., ISBN 0-471-36357-2, New Jersey, USA.

Karassik, Igor J., Messina, Joseph P., Cooper, P., Heald, C. C. (2001), *Pump Handbook*, 3rd Ed., McGraw-Hill, ISBN 0-07-034032-3, New York, USA.

Karlis, D., Xekalaki, E. (2003), *Choosing initial values for the EM algorithm for finite mixtures*, Computational Statistics and Data Analysis, Vol. 41, pp. 577-590.

Kececioglu, D. B. (2002), *Reliability & Life Testing Handbook*, Vol. 1, DEStech Publications, ISBN 1-932078-02-9, Pennsylvania, USA.

Kinaci, I., Akdogan, Y., Kus, C., Ng, H. K. T. (2014), *Statistical inference for Weibull distribution based on a modified progressive type-II censoring scheme*, Sri Lankan Journal of Applied Statistics, pp. 95-116.

Kirtay, S., Dispinar, D. (2012), *Effect of ranking selection on the Weibull modulus estimation*, Gazi University Journal of Science, Vol. 25 (1), pp. 175-187.

Kobayashi, H., Mark, B. L., Turin, W. (2012), *Probability, random processes and statistical analysis*, Cambridge University Press, ISBN 978-0-521-89544-6, Cambridge, UK.

Lawless, J.F.(2003), *Statistical Models and Methods for Lifetime Data*, Second Edition John Wiley & Sons, ISBN 0-471-37215-3, New Jersey, USA.

Lee, E. T., Wang, J. W. (2003), *Statistical methods for survival data analysis*, John Wiley & Sons, Ltd., ISBN 0-471-36997-7, New Jersey, USA.

Louit, D. M., Pascual, R., Jardine, A. K. S. (2009), *A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data*, Reliability Engineering and System Safety, Vol. 94, pp. 1618-1628.

McCool, J. I. (2012), *Using the Weibull distribution, Reliability, Modeling and Inference*, John Wiley & Sons, ISBN 978-1-118-21798-6, New York, USA.

McLachlan, G. J., Krishnan, T. (2008), *The EM algorithm and extensions*, 2nd ed., John Wiley & Sons, ISBN 978-0-471-20170-0, New Jersey, USA.

McLachlan, G. J., Krishnan, T., Ng, S. K. (2004), *The EM algorithm*, Humboldt-Universität Berlin, Center for Applied Statistics and Economics (CASE), n. 2004 (24), pp. 1-34.

McLachlan, G., Peel, D. (2000), *Finite Mixture Models*, John Wiley & Sons, ISBN 978-0-4710-0626-8, New York, USA.

Meeker, W. Q., Escobar, L. A (1998), *Statistical Methods for Reliability Data*, John Wiley & Sons, ISBN 0-471-14328-6, New Jersey, USA.

Mobley, R. K. (1999), *Root cause failure analysis*, Elsevier, ISBN 0-7506-7158-0, New York, USA.

Monchy, F. (2003), *Maintenance - Méthodes et organisations - 2e édition*, Série Gestion Industrielle, L'Usine Nouvelle, Dunod, ISBN 2-10-007816-X, Paris, França.

Montanari, G. C., Mazzanti, G., Cacciari, M., Fothergill, J. C. (1998), *Optimum estimators for the Weibull distribution from censored test data - progressively censored tests*, IEEE Transaction on Dielectrics and Electrical Insulation, Vol. 5 (2), pp. 157-164.

Montgomery, D. C., Runger, G. C. (2003), *Applied statistics and probability for engineers*, 3rd Ed, John Wiley & Sons, ISBN 0-471-20454-4, New York, USA.

Moss, T. R. (2004), *Reliability data handbook*, Professional Engineering Publishing, ISBN 978-1-86058-444-6, London, UK.

Moubray, J. (1997), *Reliability-Centered Maintenance (RCM)*, Second Edition, Industrial Press, Inc., ISBN 0-8311-3146-2, Oxford, Great Britain.

Murthy, D.N. Prabhakar (2004), *Weibull Models*, John Wiley & Sons, ISBN 0-471-36092-9, New Jersey, USA.

Nelson, W. B. (1982), *Applied Life Data Analysis*, John Wiley & Sons, ISBN 0-471-09458-7, New York, USA.

NP EN 13306 (2007), *Terminologia da Manutenção*, IPQ, Almada.

NP EN 50126 (2000), *Aplicações Ferroviárias - Especificação e demonstração de Fiabilidade, Disponibilidade, Manutibilidade e Segurança (RAMS)*, IPQ, Almada.

Ng, H. K. T., Chan, P. S., Balakrishnan, N. (2002), *Estimation of parameters from progressively censored data using EM Algorithm*, Computational Statistics and Data Analysis, Vol. 39, pp. 371-386.

O'Connor, P.D.T. & Kleyner A. (2012), *Practical Reliability Engineering - 5th Ed.*, John Wiley & Sons, Ltd., ISBN 978-0-470-97982-2, Chichester, England.

Ojile, J. O., Teixeira, J. A., Carmody, C. (2010), *Mechanical Seal Failure Analysis*, Tribology Transactions, Vol. 53 (4), pp. 630 – 635.

Olteanu, D., Freeman, L. (2010), *The evaluation of median-rank regression and maximum likelihood estimation techniques for a two parameter Weibull distribution*, Quality Engineering, Vol. 22, pp. 256-272.

Oreda (2002), *Oreda: Offshore Reliability Data Handbook - 4th Ed.*, DNV, ISBN 82-14-02705-5, Hovik, Norway.

Palgrave, R. (2003), *Troubleshooting centrifugal pumps and their systems*, Elsevier, ISBN 1-85617-391-7, Oxford, UK.

Panahi, H., Asadi, S. (2011), *Estimation of the Weibull distribution based on type II censored samples*, Applied Mathematical Sciences, Vol. 5 (52), pp. 2549-2558.

Pandis, N. (2013), *The P value problem*, American Journal of Orthodontics and Dentofacial Orthopedics, Vol. 143 (1), pp. 150-151.

Pereira, F. J. D., Sena, F. M. V. (2012), *Fiabilidade e a sua aplicação à Manutenção*, Publindústria, ISBN 978-972-8953-99-7, Porto, Portugal.

Procaccia, H., Ferton, É., Procaccia, M. (2011), *Fiabilité et maintenance des matériels industriels réparables et non réparables*, Ed. Tec & Doc, ISBN 978-2-7430-1362-2, Paris, France.

Rinne, H. (2009), *The Weibull Distribution - A Handbook*, CRC Press, ISBN 978-1-4200-8743-7, Florida, USA.

Rohde, C. A. (2014), *Introductory statistical inference with the likelihood function*, Springer, ISBN 978-3-319-10461-4, New York, USA.

Scholz, F. W. (2008), *Inference for the Weibull distribution*, Stat 498B Industrial Statistics, Maio 22, pp. 6-10.

Seber, G. A. F., Lee, A. J. (2003), *Linear Regression Analysis*, John Wiley & Sons, ISBN 0-471-41540-5, New York, USA.

Sobral, J. (2011), *Utilização da metodologia RAMS na análise de barreiras de segurança*, Tese de Doutoramento, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal.

Stanford, J. L., Vardeman, S. B. (1994), *Statistical methods for physical science*, Academic Press, ISBN 0-12-475973-4, San Diego, USA.

Taghipour, S., Banjevic, D. (2013), *Maximum likelihood estimation from interval censored recurrent event data*, Computers & Industrial Engineering, Vol. 64, pp. 143-152.

Teimouri, M., Hoseini, S. M., Nadarajah, S. (2015), *Comparison of estimation methods for the Weibull distribution*, Journal of Theoretical and Applied Statistics, Vol. 47 (1), pp. 93-109.

Tobias, P. A., Trindade, D. C. (2011), *Applied reliability* - 3th Ed, Chapman & Hall/ CRC Press, ISBN 978-1-58488-466-8, Florida, USA.

Vaurio, J. K. (1998), *Identification of process and distribution characteristics by testing monotonic and non-monotonic trends in failure intensities and hazard rates*, Reliability Engineering and System Safety, 64, 345-357.

Wang, R., Sha, N., Gu, B., Xu, X. (2014), *Statistical analysis of a Weibull extension with bathtub-shape failure rate function*, Hindawi Publishing Corporation, Advances in Statistics, Vol. 2014, pp. 1-15.

Wang, W. (2001), *Improving the quality of field data reliability analysis*, Proceedings Annual Reliability and Maintainability Symposium, IEEE Press: Piscataway, NJ, pp. 285-289.

Wang, W. (2004), *Refined rank regression method with censors*, Quality and Reliability Engineering International, Vol. 20, pp. 667-678.

Weibull, W. (1951), *A statistical distribution function of wide applicability*, Journal of Applied Mechanics, Vol. 18, pp. 293-297.

Yahaya, A. S., Yee, C. S., Ramli, N. A., Ahmad, F. (2012), *Determination of the best probability plotting position for predicting parameters of the Weibull distribution*, International Journal of Applied Science and Technology, Vol. 2 (3), pp. 106-111.

Zhang, L. F., Xie, M., Tang, L. C. (2006), *Bias correction for the least squares estimator of Weibull shape parameter with complete and censored data*, Reliability Engineering and System Safety, Vol. 91, pp. 930-939.

Zhang, L. F., Xie, M., Tang, L. C. (2007), *A study of two estimation approaches for parameters of Weibull distribution based on WPP*, Reliability Engineering and System Safety, Vol. 92, pp. 360-368.

Zio, E. (2013), *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*, Springer, ISBN 978-1-4471-4587-5, London, UK.

Anexo A

Simulação de Monte Carlo

A simulação de Monte Carlo utiliza algoritmos geradores de números aleatórios para modelar acontecimentos estocásticos.

A simulação de Monte Carlo apresenta várias vantagens na sua utilização, em comparação com simulação analítica (Dubi, 2000).

A primeira, prende-se com a capacidade da simulação de Monte Carlo de produzir uma distribuição de resultados possíveis em vez do resultado esperado apenas.

Outra vantagem é a capacidade de facilmente ser possível modelizar os parâmetros dos componentes como variáveis aleatórias características pelas funções de distribuição de probabilidades, em vez de valores constantes. Uma simulação de Monte Carlo também é capaz de mais facilmente modelar um comportamento complexo dum sistema, considerando probabilidades condicionais, falhas em cascata, acontecimentos não exclusivos (independentes), etc.

A simulação de Monte Carlo apresenta também desvantagens comparativas com a simulação analítica.

A mais significativa tem a ver com a chamada intensidade computacional. Enquanto a simulação analítica apenas necessita simular o que é expectável que se venha a passar num determinado momento, uma simulação de Monte Carlo necessita simular acontecimentos numa grande quantidade de momentos.

Outra desvantagem é a imprecisão. Mesmo utilizando um número muito grande de momentos de análise, várias simulações de Monte Carlo sobre o mesmo problema em análise darão origem a resultados ligeiramente diferentes. Esta falta de precisão inibe uma análise de sensibilidade, o cálculo de gradientes e a quantificação de pequenas modificações em sistemas de grande dimensão.

A simulação de Monte Carlo possui duas abordagens básicas, a sequencial e a não sequencial, que diferem pelo tratamento dado aos aspectos cronológicos da simulação (Dubi, 2000).

Na simulação de Monte Carlo sequencial tenta-se modelar o comportamento do sistema precisamente como ele ocorre, preservando-se as características cronológicas do sistema. Este tipo de simulação produz resultados mais reais na avaliação composta de sistemas, contudo requer mais exigência computacional devido à interação entre os estados.

Na simulação de Monte Carlo não sequencial os estados do sistema podem ser selecionados de forma probabilística e simuladas por uma ordem arbitrária. Não existe a preocupação da cronologia dos eventos. Este tipo de simulação tende a ser menos intensivo do ponto de vista computacional e as regras de simulação são mais simples uma vez que as contingências não interagem entre si.

A utilização da simulação de Monte Carlo sequencial é apropriada quando o sistema é muito dependente de acontecimentos anteriores. As aplicações que não têm esta característica podem utilizar a simulação de Monte Carlo sequencial, contudo pode-se reduzir o tempo de cálculo computacional, sem ter que sacrificar a precisão de cálculo pela utilização da simulação de Monte Carlo não sequencial.

Os resultados da simulação são frequentemente apresentados sob diferentes combinações dos parâmetros experimentais.

Para os dados completos, existem cinco fatores importantes, os valores dos parâmetros reais de η e β (denominados de η_T e β_T), o tamanho da amostra n , número de iteração M e a geração de números aleatórios. Para os dados censurados, há mais um fator, designadamente, o nível de censura c . A seguir são resumidamente apresentadas algumas considerações sobre a seleção dos valores para cada parâmetro.

i) Seleção do valor de η_T e β_T

Para a distribuição de Weibull, η é o parâmetro de escala que pode assumir qualquer valor positivo e β é o parâmetro de forma que geralmente tem valores entre 0.1 e 10.

Para ter uma análise completa do desempenho dos estimadores seria necessário examinar todas as combinações dos valores mais comuns de η_T e β_T . Contudo, conforme referido em (Rohde, 2014), (Scholz, 2008), (Wang, 2001), (Zhang, 2007) pela utilização das funções pivotais,

$$\frac{\hat{\beta}}{\beta} \tag{A.1}$$

$$\frac{\hat{\eta}}{\eta} \tag{A.2}$$

pode ser bastante reduzido o trabalho de simulação.

Os teoremas apresentados a seguir ilustram claramente as propriedades destes dois parâmetros na simplificação da complexidade da simulação.

Teorema 1: $\hat{\beta} / \beta$ é independente de η e β e tem o mesmo comportamento do que $\hat{\beta}_{1,1}$

Teorema 2: $\hat{\eta} / \eta$ depende apenas de β e tem o mesmo comportamento do que $\hat{\eta}_{1,\beta}$

Sendo que $\hat{\eta}_{1,1}$ e $\hat{\beta}_{1,1}$ denota o estimador de η e β quando a amostra é uma distribuição de Weibull normalizada, ou seja, $\eta_T = \beta_T = 1$.

O teorema 1 refere que $\hat{\beta} / \beta$ tem um comportamento semelhante ao $\hat{\beta}_{1,1}$, assim, as propriedades tais como o enviesamento, a variância e o erro quadrático médio do estimador sob qualquer combinação de η_T e β_T podem ser obtidas a partir de $\hat{\beta}_{1,1}$ ($\eta_T = \beta_T = 1$). Assim para analisar o parâmetro de forma estimado os valores de η_T e β_T podem ser fixos e iguais a 1, em toda a simulação.

Por outro lado, para analisar o parâmetro de escala estimado, com base no teorema 2, $\hat{\eta} / \eta$ tem a mesma distribuição do que $\hat{\eta}_{1,\beta}$ e depende apenas de β . O valor η_T pode ser fixo e igual a 1, no entanto devem ser utilizados diferentes valores de β_T .

Em resumo, o valor de η_T pode ser sempre igual a 1. Para analisar $\hat{\beta}$, β_T pode ser também igual a 1. No entanto, para analisar $\hat{\eta}$ vários valores de β_T devem ser considerados, por exemplo, $\beta_T = 0.5, 1, 3$.

ii) Seleção do tamanho da amostra

A seleção do tamanho da amostra depende do foco do estudo. Neste trabalho, o foco são as amostras de pequena dimensão, que também é o foco de alguns estudos (Guure, 2013), (Hossain, 2003), (Olteanu, 2010).

Para uma amostra de tamanho n , é habitualmente aceite como uma pequena amostra, se $n \leq 20$, uma amostra de tamanho médio se $20 < n \leq 100$ e uma grande amostra se $n > 100$ (Abernethy, 2006).

Com o foco em pequenas e médias amostras, o tamanho da amostra, n neste trabalho é definido no intervalo de 4 a 100.

iii) Seleção do número de iterações

A precisão dos resultados da simulação está intimamente relacionada com o número de iteração ou o número de repetições. Normalmente, o aumento do número de iteração pode conseguir uma maior precisão, no entanto, o tempo da simulação é também maior. Deve haver um compromisso entre a precisão e o tempo de simulação. A precisão dos resultados da simulação para um determinado número de iterações pode ser simplesmente estimada, repetindo todo o processo de simulação por várias vezes.

Na literatura, 10000 é o número da iteração habitualmente usado (Zio, 2013). Para se ter uma precisão mais elevada, 50000 repetições é o indicado.

iv) Geração de números aleatórios

A simulação de um modelo normalmente requer a geração de números aleatórios com os quais é possível obter as observações que obedecem a determinada distribuição de probabilidade (Dubi, 2000).

Existe vários métodos para gerar números aleatórios, dos quais se destaca o gerador proporcional (“congruential generator”), proposto por D. Y. Lehmer em 1948. O gerador proporcional dá origem a um número aleatório uniformemente distribuído, com um valor entre zero e um.

Quando a distribuição necessária é diferente da uniforme, o número aleatório uniforme é transformado noutro valor pela transformada inversa da função de distribuição acumulada.

O método da transformada inversa compreende os seguintes passos:

1. Gerar um número aleatório u da distribuição uniforme (0, 1).
2. Efetuar a atribuição.

$$x = F^{-1}(u),$$

onde F^{-1} representa a inversa da função de distribuição acumulada que se deseja gerar.

O software escolhido neste estudo para realizar a simulação de Monte Carlo é o R.

Este software permite simplificar os passos referidos anteriormente pela utilização da seguinte instrução para a distribuição de Weibull,

`rweibull (n, scale= , shape=),`

em que n representa o número de observações aleatórias.

v) Seleção do nível de censura

O nível de censura, c , é muitas vezes apresentado por percentagem. Habitualmente para $c \geq 50\%$ refere-se a uma amostra altamente censurada (Zhang, 2007).

Ambos os níveis, baixos e altos de censura são analisados neste estudo e a faixa de frequência é de 20% a 60%.

Referências

Abernethy, R. B. (2006), *The New Weibull Handbook*, 5th Ed, Robert B. Abernethy, ISBN 0-9653062-3-2, Florida, USA.

Dubi, A. (2000), *Monte Carlo Applications in Systems Engineering*, John Wiley & Sons, ISBN 0-471-98172-9, Chichester, UK.

Guure, C. B., Ibrahim, N. A. (2013), *Methods for estimating the 2-parameter Weibull distribution with type-I censored data*, Journal of Applied Sciences, Engineering and Technology Vol. 5 (3), pp 689-694.

Hossain, A. M., Zimmer, W. J. (2003), *Comparisons of estimation methods for Weibull parameters: complete and censored samples*, Journal of Statistical Computation and Simulation, Vol. 73 (2), pp. 145-153.

Olteanu, D., Freeman, L. (2010), *The evaluation of median-rank regression and maximum likelihood estimation techniques for a two parameter Weibull distribution*, Quality Engineering, Vol. 22, pp. 256-272.

Rohde, C. A. (2014), *Introductory statistical inference with the likelihood function*, Springer, ISBN 978-3-319-10461-4, New York, USA.

Scholz, F. W. (2008), *Inference for the Weibull distribution*, Stat 498B Industrial Statistics, Maio 22, pp. 6-10.

Wang, W. (2001), *Improving the quality of field data reliability analysis*, Proceedings Annual Reliability and Maintainability Symposium, EEE Press: Piscataway, NJ, pp. 285-289.

Zhang, L. F., Xie, M., Tang, L. C. (2007), *A study of two estimation approaches for parameters of Weibull distribution based on WPP*, Reliability Engineering and System Safety, Vol. 92, pp. 360-368.

Zio, E. (2013), *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*, Springer, ISBN 978-1-4471-4587-5, London, UK.

Anexo B

Valor esperado

O conceito de valor esperado é importante para o cálculo dos parâmetros de uma distribuição.

O valor esperado de uma função $g(x)$ da variável aleatória X é definido por:

i) no caso discreto,

$$E[g(x)] = \sum_{i=1}^n g(x_i) f(x_i) \quad \text{B.1}$$

ii) no caso contínuo,

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x) f(x) dx \quad \text{B.2}$$

Algumas propriedades do valor esperado:

i) para qualquer valor real a e b ,

$$E(a.X + b) = a.E(x) + b \quad \text{B.3}$$

$$E(a) = a \quad \text{B.4}$$

ii) para qualquer variável X e Y ,

$$E(X + Y) = E(X) + E(Y) \quad \text{B.5}$$

Se X e Y forem independentes,

$$E(X.Y) = E(X).E(Y) \quad \text{B.6}$$

O valor esperado condicional de Y dado $X = x$,

$$E(Y|X = x) = \int y.f(y|x) dy \quad \text{B.7}$$

Anexo C

Valores críticos de Kolmogorov-Smirnov

| Dimensão da amostra (<i>n</i>) | Nível de significância (α) | | | | |
|----------------------------------|-------------------------------------|-------|-------|-------|-------|
| | 0,20 | 0,10 | 0,05 | 0,02 | 0,01 |
| 1 | 0,900 | 0,950 | 0,975 | 0,990 | 0,995 |
| 2 | 0,684 | 0,776 | 0,842 | 0,900 | 0,929 |
| 3 | 0,565 | 0,636 | 0,624 | 0,689 | 0,734 |
| 4 | 0,493 | 0,565 | 0,624 | 0,689 | 0,734 |
| 5 | 0,447 | 0,509 | 0,563 | 0,627 | 0,669 |
| 6 | 0,410 | 0,468 | 0,519 | 0,577 | 0,617 |
| 7 | 0,381 | 0,436 | 0,483 | 0,538 | 0,576 |
| 8 | 0,358 | 0,410 | 0,454 | 0,507 | 0,542 |
| 9 | 0,339 | 0,387 | 0,430 | 0,480 | 0,513 |
| 10 | 0,323 | 0,369 | 0,409 | 0,457 | 0,489 |
| 11 | 0,308 | 0,352 | 0,391 | 0,437 | 0,468 |
| 12 | 0,296 | 0,338 | 0,375 | 0,419 | 0,449 |
| 13 | 0,285 | 0,325 | 0,361 | 0,404 | 0,432 |
| 14 | 0,275 | 0,314 | 0,349 | 0,390 | 0,418 |
| 15 | 0,266 | 0,304 | 0,338 | 0,377 | 0,404 |
| 16 | 0,258 | 0,295 | 0,327 | 0,366 | 0,392 |
| 17 | 0,250 | 0,286 | 0,318 | 0,355 | 0,381 |
| 18 | 0,244 | 0,279 | 0,309 | 0,346 | 0,371 |
| 19 | 0,237 | 0,271 | 0,301 | 0,337 | 0,361 |
| 20 | 0,232 | 0,265 | 0,294 | 0,329 | 0,352 |
| 21 | 0,226 | 0,259 | 0,287 | 0,321 | 0,344 |
| 22 | 0,221 | 0,253 | 0,281 | 0,314 | 0,337 |
| 23 | 0,216 | 0,247 | 0,275 | 0,307 | 0,330 |
| 24 | 0,212 | 0,242 | 0,269 | 0,301 | 0,323 |

| | | | | | |
|----|-------|-------|-------|-------|-------|
| 25 | 0,208 | 0,238 | 0,264 | 0,295 | 0,317 |
| 26 | 0,204 | 0,233 | 0,259 | 0,290 | 0,311 |
| 27 | 0,200 | 0,229 | 0,254 | 0,284 | 0,305 |
| 28 | 0,197 | 0,225 | 0,250 | 0,279 | 0,300 |
| 29 | 0,193 | 0,221 | 0,346 | 0,275 | 0,295 |
| 30 | 0,190 | 0,218 | 0,242 | 0,270 | 0,290 |
| 31 | 0,187 | 0,214 | 0,238 | 0,266 | 0,285 |
| 32 | 0,184 | 0,211 | 0,234 | 0,262 | 0,281 |
| 33 | 0,182 | 0,208 | 0,231 | 0,258 | 0,277 |
| 34 | 0,179 | 0,205 | 0,227 | 0,254 | 0,273 |
| 35 | 0,177 | 0,202 | 0,224 | 0,251 | 0,269 |
| 36 | 0,174 | 0,199 | 0,221 | 0,247 | 0,265 |
| 37 | 0,172 | 0,196 | 0,218 | 0,244 | 0,262 |
| 38 | 0,170 | 0,194 | 0,215 | 0,241 | 0,258 |
| 39 | 0,168 | 0,191 | 0,213 | 0,238 | 0,255 |
| 40 | 0,165 | 0,189 | 0,210 | 0,235 | 0,252 |